(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) **International Patent Classification**[7]: **C12N 15/10**, C12Q 1/68, A61K 31/7105

(21) **International Application Number:** PCT/US01/08435

(22) **International Filing Date:** 16 March 2001 (16.03.2001)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
60/189,739   16 March 2000 (16.03.2000)   US
60/243,097   24 October 2000 (24.10.2000)   US

(71) **Applicants** *(for all designated States except US):* **GE-NETICA, INC.** [US/US]; Building 600, One Kendall Square, Cambridge, MA 02139 (US). **COLD SPRING HARBOR LABORATORY** [US/US]; One Bungtown Road, Cold Spring Harbor, NY 11724 (US).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only):* **BEACH, David** [US/US]; 19 Woodland Drive, Huntington Bay, NY 11743 (US). **BERNSTEIN, Emily** [US/US]; 5 Kingsley Road, Huntington, NY 11743 (US). **CAUDY, Amy** [US/US]; 823 Avalon Court Drive, Melville, NY 11747 (US). **HAMMOND, Scott** [US/US]; 146 Meadowlawn Road, Huntington, NY 11743 (US). **HANNON, Gregory** [US/US]; 92 Sammis Street, Huntington, NY 11743 (US).

(74) **Agents: VINCENT, Matthew, P.** et al.; Ropes & Gray, One International Place, Boston, MA 02110 (US).

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** METHODS AND COMPOSITIONS FOR RNA INTERFERENCE

(57) **Abstract:** The present invention provides methods for attenuating gene expression in a cell using gene-targeted double stranded RNA (dsRNA). The dsRNA contains a nucleotide sequence that hybridizes under physiologic conditions of the cell to the nucleotide sequence of at least a portion of the gene to be inhibited (the "target" gene).

Methods and Compositions for RNA Interference

Government Support

Background of the Invention

"RNA interference", "post-transcriptional gene silencing", "quelling" — these
different names describe similar effects that result from the overexpression or
10 misexpression of transgenes, or from the deliberate introduction of double-stranded RNA
into cells (reviewed in Fire A (1999) Trends Genet 15:358–363; Sharp PA (1999) Genes
Dev 13:139–141; Hunter C (1999) Curr Biol 9:R440–R442; Baulcombe DC (1999) Curr
Biol 9:R599–R601; Vaucheret et al. (1998) Plant J 16:651–659). The injection of double-
stranded RNA into the nematode *Caenorhabditis elegans*, for example, acts systemically
15 to cause the post-transcriptional depletion of the homologous endogenous RNA (Fire et al.
(1998) Nature 391: 806–811; and Montgomery et al. (1998) PNAS 95:15502–15507).
RNA interference, commonly referred to as RNAi, offers a way of specifically and
potently inactivating a cloned gene, and is proving a powerful tool for investigating gene
function. But the phenomenon is interesting in its own right; the mechanism has been
20 rather mysterious, but recent research — the latest reported by Smardon et al. (2000) Curr
Biol 10:169–178— is beginning to shed light on the nature and evolution of the biological
processes that underlie RNAi.

RNAi was discovered when researchers attempting to use the antisense RNA
approach to inactivate a *C. elegans* gene found that injection of sense-strand RNA was
25 actually as effective as the antisense RNA at inhibiting gene function. Guo et al. (1995)
Cell 81:611–620. Further investigation revealed that the active agent was modest amounts
of double-stranded RNA that contaminate *in vitro* RNA preparations. Researchers quickly
determined the 'rules' and effects of RNAi. Exon sequences are required, whereas introns
and promoter sequences, while ineffective, do not appear to compromise RNAi (though
30 there may be gene-specific exceptions to this rule). RNAi acts systemically — injection
into one tissue inhibits gene function in cells throughout the animal. The results of a
variety of experiments, in *C. elegans* and other organisms, indicate that RNAi acts to
destabilize cellular RNA after RNA processing.

The potency of RNAi inspired Timmons and Fire (1998 Nature **395**: 854) to do a simple experiment that produced an astonishing result. They fed to nematodes bacteria that had been engineered to express double-stranded RNA corresponding to the *C. elegans unc-22* gene. Amazingly, these nematodes developed a phenotype similar to that of *unc-22*

5   mutants that was dependent on their food source. The ability to conditionally expose large numbers of nematodes to gene-specific double-stranded RNA formed the basis for a very powerful screen to select for RNAi-defective *C. elegans* mutants and then to identify the corresponding genes.

Double-stranded RNAs (dsRNAs) can provoke gene silencing in numerous in vivo

10   contexts including *Drosophila, Caenorhabditis elegans,* planaria, hydra, trypanosomes, fungi and plants.   However, the ability to recapitulate this phenomenon in higher eukaryotes, particularly mammalian cells, has not be accomplished in the art. Nor has the prior art demonstrated that this phenomena can be observe in cultured eukaryotes cells.

15   **Summary of the Invention**

One aspect of the present invention provides a method for attenuating expression of a target gene in a non-embryonic cell suspended in culture, comprising introducing into the cell a double stranded RNA (dsRNA) in an amount sufficient to attenuate expression of the target gene, wherein the dsRNA comprises a nucleotide sequence that hybridizes

20   under stringent conditions to a nucleotide sequence of the target gene.

Another aspect of the present invention provides a method for attenuating expression of a target gene in a mammalian cell, comprising

(i)   activating one or both of a Dicer activity or an Argonaut activity in the cell, and ·

25   (ii)   introducing into the cell a double stranded RNA (dsRNA) in an amount sufficient to attenuate expression of the target gene, wherein the dsRNA comprises a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence of the target gene.

In certain embodiments, the cell is suspended in culture; while in other embodiments the

30   cell is in a whole animal, such as a non-human mammal.

In certain preferred embodiments, the cell is engineered with (i) a recombinant gene encoding a Dicer activity, (ii) a recombinant gene encoding an Argonaut activity, or (iii) both. For instance, the recombinant gene may encode, for a example, a protein which includes an amino acid sequence at least 50 percent identical to SEQ ID No. 2 or 4; or be

defined by a coding sequence hybridizes under wash conditions of 2 x SSC at 22°C to SEQ ID No. 1 or 3. In certain embodiments, the recombinant gene may encode, for a example, a protein which includes an amino acid sequence at least 50 percent identical to the Argonaut sequence shown in Figure 24.

5   In certain embodiments, rather than use a heterologous expression construct(s), an endogenous Dicer gene or Argonaut gene can be activated, e.g, by gene activation technology, expression of activated transcription factors or other signal transduction protein, which induces expression of the gene, or by treatment with an endogenous factor which upregualtes the level of expression of the protein or inhibits the degradation of the 10 protein.

In certain preferred embodiments, the target gene is an endogenous gene of the cell. In other embodiments, the target gene is an heterologous gene relative to the genome of the cell, such as a pathogen gene, e.g., a viral gene.

In certain embodiments, the cell is treated with an agent that inhibits protein kinase 15 RNA-activated (PKR) apoptosis, such as by treatment with agents which inhibit expression of PKR, cause its destruction, and/or inhibit the kinase activity of PKF.

In certain preferred embodiments, the cell is a primate cell, such as a human cell.

In certain embodiments, the dsRNA is at least 50 nucleotides in length, and preferably 400-800 nucleotides in length.

20   Still another aspect of the present invention provides an assay for identifying nucleic acid sequences responsible for conferring a particular phenotype in a cell, comprising

(i)   constructing a variegated library of nucleic acid sequences from a cell in an orientation relative to a promoter to produce double stranded DNA;

25   (ii)   introducing the variegated dsRNA library into a culture of target cells, which cells have an activated Dicer activity or Argonaut activity;

(iii)   identifying members of the library which confer a particular phenotype on the cell, and identifying the sequence from a cell which correspond, such as being identical or homologous, to the library member.

30

Yet another aspect of the present invention provides a method of conducting a drug discovery business comprising:

(i)   identifying, by the assay of claim 16, a target gene which provides a phenotypically desirable response when inhibited by RNAi;

(ii)   identifying agents by their ability to inhibit expression of the target gene or the activity of an expression product of the target gene;

(iii)   conducting therapeutic profiling of agents identified in step (b), or further analogs thereof, for efficacy and toxicity in animals; and

(iv)   formulating a pharmaceutical preparation including one or more agents identified in step (iii) as having an acceptable therapeutic profile.

The method may include an additional step of establishing a distribution system for distributing the pharmaceutical preparation for sale, and may optionally include establishing a sales group for marketing the pharmaceutical preparation.

Another aspect of the present invention provides a method of conducting a target discovery business comprising:

(i)   identifying, by the assay of claim 16, a target gene which provides a phenotypically desirable response when inhibited by RNAi;

(ii)   (optionally) conducting therapeutic profiling of the target gene for efficacy and toxicity in animals; and

(iii).   licensing, to a third party, the rights for further drug development of inhibitors of the target gene.

Another aspect of the invention provides a method for inhibiting RNAi by inhibiting the expression or activity of an RNAi enzyme. Thus, the subject method may include inhibiting the acitivity of Dicer and/or the 22-mer RNA.

Still another aspect relates to the a method for altering the specificity of an RNAi by modifying the sequence of the RNA component of the RNAi enzyme.

Another aspect of the invention relates to purified or semi-purified preparations of the RNAi enzyme or components thereof. In certain embodiments, the preparations are used for identifying compounds, especially small organic molecules, which inhibit or potentiate the RNAi activity. Small molecule inhibitors, for example, can be used to inhibit dsRNA responses in cells which are purposefully being transfected with a virus which produces double stranded RNA.

The dsRNA construct may comprise one or more strands of polymerized ribonucleotide. It may include modifications to either the phosphate-sugar backbone or the nucleoside. The double-stranded structure may be formed by a single self-complementary

RNA strand or two complementary RNA strands. RNA duplex formation may be initiated either inside or outside the cell. The dsRNA construct may be introduced in an amount which allows delivery of at least one copy per cell. Higher doses of double-stranded material may yield more effective inhibition. Inhibition is sequence-specific in that

5   nucleotide sequences corresponding to the duplex region of the RNA are targeted for genetic inhibition. dsRNA constructs containing a nucleotide sequences identical to a portion of the target gene is preferred for inhibition. RNA sequences with insertions, deletions, and single point mutations relative to the target sequence have also been found to be effective for inhibition. Thus, sequence identity may optimized by alignment

10   algorithms known in the art and calculating the percent difference between the nucleotide sequences. Alternatively, the duplex region of the RNA may be defined functionally as a nucleotide sequence that is capable of hybridizing with a portion of the target gene transcript.

15   **Brief Description of the Drawings**

Figure 1: RNAi in S2 cells. a, *Drosophila* S2 cells were transfected with a plasmid that directs *lacZ* expression from the copia promoter in combination with dsRNAs corresponding to either human CD8 or lacZ, or with no dsRNA, as indicated. b, S2 cells were co-transfected with a plasmid that directs expression of a GFP–US9 fusion protein

20   (12) and dsRNAs of either *lacZ* or *cyclin E,* as indicated. Upper panels show FACS profiles of the bulk population. Lower panels show FACS profiles from GFP-positive cells. c, Total RNA was extracted from cells transfected with *lacZ, cyclin E, fizzy* or *cyclin A* dsRNAs, as indicated. Northern blots were hybridized with sequences not present in the transfected dsRNAs.

25   Figure 2: RNAi *in vitro*. a, Transcripts corresponding to either the first 600 nucleotides of *Drosophila cyclin E* (E600) or the first 800 nucleotides of *lacZ* (Z800) were incubated in lysates derived from cells that had been transfected with either *lacZ* or *cyclin E* (cycE) dsRNAs, as indicated. Time points were 0, 10, 20, 30, 40 and 60 min for *cyclin E* and 0, 10, 20, 30 and 60 min for *lacZ*. b, Transcripts were incubated in an extract of S2

30   cells that had been transfected with *cyclin E* dsRNA (cross-hatched box, below). Transcripts corresponded to the first 800 nucleotides of *lacZ* or the first 600, 300, 220 or 100 nucleotides of *cyclin E,* as indicated. Eout is a transcript derived from the portion of the *cyclin E* cDNA not contained within the transfected dsRNA. E-ds is identical to the dsRNA that had been transfected into S2 cells. Time points were 0 and 30 min.

35   c, Synthetic transcripts complementary to the complete *cyclin E* cDNA (Eas) or the final

600 nucleotides (Eas600) or 300 nucleotides (Eas300) were incubated in extract for 0 or 30 min.

Figure 3: Substrate requirements of the RISC. Extracts were prepared from cells transfected with *cyclin E* dsRNA. Aliquots were incubated for 30 min at 30 °C before the
5   addition of either the *cyclin E* (E600) or *lacZ* (Z800) substrate. Individual 20-μl aliquots, as indicated, were pre-incubated with 1 mM $CaCl_2$ and 5 mM EGTA, 1 mM $CaCl_2$, 5 mM EGTA and 60 U of micrococcal nuclease, 1 mM $CaCl_2$ and 60 U of micrococcal nuclease or 10 U of DNase I (Promega) and 5 mM EGTA. After the 30-min pre-incubation, EGTA was added to those samples that lacked it. Yeast tRNA (1 μg) was added to all samples.
10   Time points were at 0 and 30 min.

Figure 4: The RISC contains a potential guide RNA. a, Northern blots of RNA from either a crude lysate or the S100 fraction (containing the soluble nuclease activity, see Methods) were hybridized to a riboprobe derived from the sense strand of the *cyclin E* mRNA. b, Soluble *cyclin-E*-specific nuclease activity was fractionated as described in
15   Methods. Fractions from the anion-exchange resin were incubated with the lacZ, control substrate (upper panel) or the cyclin E substrate (centre panel). Lower panel, RNA from each fraction was analysed by northern blotting with a uniformly labelled transcript derived from sense strand of the *cyclin E* cDNA. DNA oligonucleotides were used as size markers.

20   Figure 5: Generation of 22mers and degradation of mRNA are carried out by distinct enzymatic complexes. A. Extracts prepared either from 0-12 hour *Drosophila* embryos or *Drosophila* S2 cells (see Methods) were incubated 0, 15, 30, or 60 minutes (left to right) with a uniformly-labeled double-stranded RNA corresponding to the first 500 nucleotides of the *Drosophila cyclin E* coding region. M indicates a marker prepared
25   by *in vitro* transcription of a synthetic template. The template was designed to yield a 22 nucleotide transcript. The doublet most probably results from improper initiation at the +1 position. B. Whole-cell extracts were prepared from S2 cells that had been transfected with a dsRNA corresponding to the first 500 nt. of the luciferase coding region. S10 extracts were spun at 30,000xg for 20 minutes which represents our standard RISC
30   extract[6]. S100 extracts were prepared by further centrifugation of S10 extracts for 60 minutes at 100,000xg. Assays for mRNA degradation were carried out as described previously[6] for 0,30 or 60 minutes (left to right in each set) with either a single-stranded luciferase mRNA or a single-stranded cyclin E mRNA, as indicated. C. S10 or S100 extracts were incubated with cyclin E dsRNAs for 0, 60 or 120 minutes (L to R).

35   Figure 6: Production of 22mers by recombinant CG4792/Dicer. A. *Drosophila* S2 cells were transfected with plasmids that direct the expression of T7-epitope tagged

versions of Drosha, CG4792/Dicer-1 and Homeless. Tagged proteins were purified from cell lysates by immunoprecipitation and were incubated with *cyclin E* dsRNA. For comparison, reactions were also performed in *Drosophila* embryo and S2 cell extracts. As a negative control, immunoprecipitates were prepared from cells transfected with a β-
5   galactosidase expression vector. Pairs of lanes show reactions performed for 0 or 60 minutes. The synthetic marker (M) is as described in the legend to Figure 1. B. Diagrammatic representations of the domain structures of CG4792/Dicer-1, Drosha and Homeless are shown. C. Immunoprecipitates were prepared from detergent lysates of S2 cells using an antiserum raised against the C-terminal 8 amino acids of Drosophila Dicer-1
10   (CG4792). As controls, similar preparations were made with a pre-immune serum and with an immune serum that had been pre-incubated with an excess of antigenic peptide. Cleavage reactions in which each of these precipitates was incubated with an ~500 nt. fragment of Drosophila cyclin E are shown. For comparsion, an incubation of the substrate in Drosophila embryo extract was electrophoresed in parallel. D. Dicer
15   immunoprecipitates were incubated with dsRNA substrates in the presence or absence of ATP. For comparison, the same substrate was incubated with S2 extracts that either contained added ATP or that were depleted of ATP using glucose and hexokinase (see methods). E. Drosophila S2 cells were transfected with uniformly, 32P-labelled dsRNA corresponding to the first 500 nt. of GFP. RISC complex was affinity purified using a
20   histidine-tagged version of D.m. Ago-2, a recently identified component of the RISC complex (Hammond et al., in prep). RISC was isolated either under conditions in which it remains ribosome associated (ls, low salt) or under conditions that extract it from the ribosome in a soluble form (hs, high salt)[6]. For comparison, the spectrum of labelled RNAs in the total lysate is shown. F. Guide RNAs produced by incubation of dsRNA
25   with a Dicer immunoprecipitate are compared to guide RNAs present in a affinity-purified RISC complex. These precisely comigrate on a gel that has single-nucleotide resolution. The lane labelled control is an affinity selection for RISC from cell that had been transfected with labeled dsRNA but not with the epitope-tagged D.m. Ago-2.

Figure 7: Dicer participates in RNAi. A. Drosophila S2 cells were transfected
30   with dsRNAs corresponding to the two Drosophila Dicers (CG4792 and CG6493) or with a control dsRNA corresponding to murine caspase 9. Cytoplasmic extracts of these cells were tested for Dicer activity. Transfection with Dicer dsRNA reduced activity in lysates by 7.4-fold. B. The Dicer-1 antiserum (CG4792) was used to prepare immunoprecipitates from S2 cells that had been treated as described above. Dicer dsRNA reduced the activity
35   of Dicer-1 in this assay by 6.2-fold. C. Cells that had been transfected two days previously with either mouse caspase 9 dsRNA or with Dicer dsRNA were cotransfected with a GFP expression plasmid and either control, luciferase dsRNA or GFP dsRNA.

Three independent experiments were quantified by FACS. A comparison of the relative percentage of GFP-positive cells is shown for control (GFP plasmid plus luciferase dsRNA) or silenced (GFP plamsid plus GFP dsRNA) populations in cells that had previously been transfected with either control (caspase 9) or Dicer dsRNAs.

5   Figure 8: Dicer is an evolutionarily conserved ribonuclease. A. A model for production of 22mers by Dicer. Based upon the proposed mechanism of action of Ribonuclease III, we propose that Dicer acts on its substrate as a dimer. The positioning of the two ribonuclease domains (RIIIa and RIIIb) within the enzyme would thus determine the size of the cleavage product. An equally plausible alternative model could
10   be derived in which the RIIIa and RIIIb domains of each Dicer enzyme would cleave in concert at a single position. In this model, the size of the cleavage product would be determined by interaction between two neighboring Dicer enzymes. B. Comparison of the domain structures of potential Dicer homologs in various organisms (*Drosophila* - CG4792, CG6493, *C. elegans* - K12H4.8, *Arabidopsis* – CARPEL FACTORY[24],
15   T25K16.4, AC012328_1, human Helicase-MOI[25] and *S. pombe* - YC9A_SCHPO). The ZAP domains were identified both by analysis of individual sequences with Pfam[27] and by Psi-blast[28] searches. The ZAP domain in the putative *S. pombe* Dicer is not detected by PFAM but is identified by Psi-Blast and is thus shown in a different color. For comparison, a domain structure of the RDE1/QDE2/ARGONAUTE family is shown. It
20   should be noted that the ZAP domains are more similar within each of the Dicer and ARGONAUTE families than they are between the two groups. C. An alignment of the ZAP domains in selected Dicer and Argonaute family members is shown. The alignment was produced using ClustalW.

Figure 9: Purification strategy for RISC. (second step in RNAi model).

25   Figure 10: Fractionation of RISC activity over sizing column. Actvity fractionates as 500KD complex. Also, antibody to dm argonaute 2 cofractionates with activity.

Figure 11-13: Fractionation of RISC over monoS, monoQ, Hydroxyapatite columns. Dm argonaute 2 protein also cofactionates.

Figure 14: Alignment of dm argonaute 2 with other family members.

30   Figure 15: Confirmation of dm argonaute 2. S2 cells were transfected with labeled dsRNA and His tagged argonaute. Argonaute was isolated on nickel agarose and RNA component was identified on 15% acrylamide gel.

Figure 16: S2 cell and embryo extracts were assayed for 22mer generating activity.

Figure 17: RISC can be separated from 22mer generating activity (dicer). Spinning extracts (S100) can clear RISC activity from supernatant (left panel) however, S100 spins still contain dicer activity (right panel).

Figure 18: Dicer is specific for dsRNA and prefers longer substrates.

Figure 19: Dicer was fractionated over several columns.

Figure 20: Identification of dicer as enzyme which can process dsRNA into 22mers. Various RNaseIII family members were expressed with n terminal tags, immunoprecipitated, and assayed for 22mer generating activity (left panel). In right panel, antibodies to dicer could also precipitate 22mer generating activity.

Figure 21: Dicer requires ATP.

Figure 22: Dicer produces RNAs that are the same size as RNAs present in RISC.

Figure 23: Human dicer homolog when expressed and immunoprecipitated has 22mer generating activity.

Figure 24: Sequence of dm argonaute 2. Peptides identified by microsequencing are shown in underline.

Figure 25: Molecular charaterization of dm argonaute 2. The presence of an intron in coding sequence was determined by northern blotting using intron probe. This results in a different 5' reading frame that that published genome seqeunce. Number of polyglutaine repeats was determined by genomic PCR.

Figure 26: Dicer activity can be created in human cells by expression of human dicer gene. Host cell was 293. Crude extracts had dicer activity, while activity was absent from untransfected cells. Activity is not dissimilar to that seen in drosophila embryo extracts..

Figure 27: An ~500 nt. fragment of the gene that is to be silenced (X) is inserted into the modified vector as a stable direct repeat using standard cloning procedures. Treatment with commercially available cre recombinase reverses sequences within the loxP sites (L) to create an inverted repeat. This can be stably maintained and amplified in an sbc mutant bacterial strain (DL759). Transcription in vivo from the promoter of choice (P) yields a hairpin RNA that causes silencing. A zeocin resistance marker is included to insure maintenance of the direct and inverted repeat structures; however this is non-essential in vivo and could be removed by pre-mRNA splicing if desired. Smith, N. A. *et al.* Total silencing by intron-spliced hairpin RNAs. *Nature* 407, 319-20 (2000).

Detailed Description of the Certain Preferred Embodiments

I. Overview

The present invention provides methods for attenuating gene expression in a cell using gene-targeted double stranded RNA (dsRNA). The dsRNA contains a nucleotide sequence that hybridizes under physiologic conditions of the cell to the nucleotide sequence of at least a portion of the gene to be inhibited (the "target" gene).

A significant aspect to certain embodiments of the present invention relates to the demonstration in the present application that RNAi can in fact be accomplished in cultured cells, rather than whole organisms as decribed in the art.

Another salient feature of the present invention concerns the ability to carry out RNAi in higher eukaryotes, particularly in non-oocytic cells of mammals, e.g., cells from adult mammals as an example.

As described in further detail below, the present invention(s) are based on the discovery that the RNAi phenomenum is mediated by a set of enzyme activities, including an essential RNA component, that are evolutionarily conserved in eukaryotes ranging from plants to mammals.

One enzyme contains an essential RNA component. After partial purification, a multi-component nuclease (herein "RISC nuclease") co-fractionates with a discrete, 22-nucleotide RNA species which may confer specificity to the nuclease through homology to the substrate mRNAs. The short RNA molecules are generated by a processing reaction from the longer input dsRNA. Without wishing to be bound by any particular theory, these 22mer guide RNAs may serve as guide sequences that instruct the RISC nuclease to destroy specific mRNAs corresponding to the dsRNA sequences.

The appended examples also identify an enzyme, Dicer, that can produce the putative guide RNAs. Dicer is a member of the RNAse III family of nucleases that specifically cleave dsRNA and is evolutionarily conserved in worms, flies, plants, fungi and, as described herein, mammals. The enzyme has a distinctive structure which includes a helicase domain and dual RNAse III motifs. Dicer also contains a region of homology to the RDE1/QDE2/ARGONAUTE family, which have been genetically linked to RNAi in lower eukaryotes. Indeed, activation of, or overexpression of Dicer may be sufficient in many cases to permit RNA interference in otherwise non-receptive cells, such as cultured eukaryotic cells, or mammalian (non-oocytic) cells in culture or in whole organisms.

In certain embodiments, the cells can be treated with an agent(s) that inhibits the double-stranded RNA-dependent protein known as PKR (protein kinase RNA-activated). Double stranded RNAs in mammalian cells typically activate protein kinase PKR that phosphorylates and inactivates eIF2a (Fire (1999) Trends Genet 15:358). The ensuing inhibition of protein synthesis ultimately results in apoptosis. This sequence-independent response may reflect a form of primitive immune response, since the presence of dsRNA is a common feature of many viral lifecycles. However, as described herein, Applicants have demonstrated that the PKR response can be overcome in favor of the sequence-specific RNAi response. However, in certain instances, it can be desirable to treat the cells with agents which inhibit expression of PKR, cause its destruction, and/or inhibit the kinase activity of PKF are specifically contemplated for use in the present method. Likewise, overexpression of or agents which ectopic activate IF2α can be used.

Thus, the present invention provides a process and compositions for inhibiting expression of a target gene in a cell, expecially a mammalian cell. In certain embodiments, the process comprises introduction of RNA (the "dsRNA construct") with partial or fully double-stranded character into the cell or into the extracellular environment. Inhibition is specific in that a nucleotide sequence from a portion of the target gene is chosen to produce the dsRNA construct. In preferred embodiments, the method utilizes a cell in which Dicer and/or Argonaute activities are recombinantly expressed or otherwise ectopically activated. This process can be (1) effective in attenuating gene expression, (2) specific to the targeted gene, and (3) general in allowing inhibition of many different types of target gene.

II. *Definitions*

For convenience, certain terms employed in the specification, examples, and appended claims are collected here.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to that it has been linked. One type of vector is a genomic integrated vector, or "integrated vector", which can become integrated into the chromsomal DNA of the host cell. Another type of vector is an episomal vector, i.e., a nucleic acid capable of extra-chromosomal replication. Vectors capable of directing the expression of genes to that they are operatively linked are referred to herein as "expression vectors". In the present specification, "plasmid" and "vector" are used interchangeably unless otherwise clear from the context.

As used herein, the term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as applicable to the embodiment being described, single-stranded (such as sense or antisense) and double-stranded polynucleotides.

5   As used herein, the term "gene" or "recombinant gene" refers to a nucleic acid comprising an open reading frame encoding a polypeptide of the present invention, including both exon and (optionally) intron sequences. A "recombinant gene" refers to nucleic acid encoding such regulatory polypeptides, that may optionally include intron sequences that are derived from chromosomal DNA. The term "intron" refers to a DNA

10   sequence present in a given gene that is not translated into protein and is generally found between exons. As used herein, the term "transfection" means the introduction of a nucleic acid, e.g., an expression vector, into a recipient cell by nucleic acid-mediated gene transfer.

A "protein coding sequence" or a sequence that "encodes" a particular polypeptide

15   or peptide, is a nucleic acid sequence that is transcribed (in the case of DNA) and is translated (in the case of mRNA) into a polypeptide in vitro or in vivo when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but is not limited to, cDNA

20   from procaryotic or eukaryotic mRNA, genomic DNA sequences from procaryotic or eukaryotic DNA, and even synthetic DNA sequences. A transcription termination sequence will usually be located 3' to the coding sequence.

Likewise, "encodes", unless evident from its context, will be meant to include DNA sequences that encode a polypeptide, as the term is typically used, as well as DNA

25   sequences that are transcribed into inhibitory antisense molecules.

The term "loss-of-function", as it refers to genes inhibited by the subject RNAi method, refers a diminishment in the level of expression of a gene when compared to the level in the absense of dsRNA constructs.

The term "expression" with respect to a gene sequence refers to transcription of the

30   gene and, as appropriate, translation of the resulting mRNA transcript to a protein. Thus, as will be clear from the context, expression of a protein coding sequence results from transcription and translation of the coding sequence.

"Cells," "host cells" or "recombinant host cells" are terms used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to

35   the progeny or potential progeny of such a cell. Because certain modifications may occur

in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

By "recombinant virus" is meant a virus that has been genetically altered, e.g., by 5   the addition or insertion of a heterologous nucleic acid construct into the particle.

As used herein, the terms "transduction" and "transfection" are art recognized and mean the introduction of a nucleic acid, e.g., an expression vector, into a recipient cell by nucleic acid-mediated gene transfer. "Transformation", as used herein, refers to a process in which a cell's genotype is changed as a result of the cellular uptake of exogenous DNA 10 or RNA, and, for example, the transformed cell expresses a dsRNA contruct.

"Transient transfection" refers to cases where exogenous DNA does not integrate into the genome of a transfected cell, e.g., where episomal DNA is transcribed into mRNA and translated into protein.

A cell has been "stably transfected" with a nucleic acid construct when the nucleic 15 acid construct is capable of being inherited by daughter cells.

As used herein, a "reporter gene construct" is a nucleic acid that includes a "reporter gene" operatively linked to at least one transcriptional regulatory sequence. Transcription of the reporter gene is controlled by these sequences to which they are linked. The activity of at least one or more of these control sequences can be directly or 20 indirectly regulated by the target receptor protein. Exemplary transcriptional control sequences are promoter sequences. A reporter gene is meant to include a promoter-reporter gene construct that is heterologously expressed in a cell.

As used herein, "transformed cells" refers to cells that have spontaneously converted to a state of unrestrained growth, i.e., they have acquired the ability to grow 25 through an indefinite number of divisions in culture. Transformed cells may be characterized by such terms as neoplastic, anaplastic and/or hyperplastic, with respect to their loss of growth control. For purposes of this invention, the terms "transformed phenotype of malignant mammalian cells" and "transformed phenotype " are intended to encompass, but not be limited to, any of the following phenotypic traits associated with 30 cellular transformation of mammalian cells: immortalization, morphological or growth transformation, and tumorigenicity, as detected by prolonged growth in cell culture, growth in semi-solid media, or tumorigenic growth in immuno-incompetent or syngeneic animals.

As used herein, "proliferating" and "proliferation" refer to cells undergoing 35 mitosis.

- 13 -

As used herein, "immortalized cells" refers to cells that have been altered via chemical, genetic, and/or recombinant means such that the cells have the ability to grow through an indefinite number of divisions in culture.

The "growth state" of a cell refers to the rate of proliferation of the cell and the
5   state of differentiation of the cell.

III. _Exemplary embodiments of Isolation Method_

One aspect of the invention provides a method for potentiating RNAi by induction or ectopic activation of an RNAi enzyme in a cell (in vivo or in vitro) or cell-free
10   mixtures. In preferred embodiments, the RNAi activity is activated or added to a mammalian cell, e.g., a human cell, which cell may be provided in vitro or as part of a whole organism. In other embodiments, the subject method is carried out using eukaryotic cells generally (except for oocytes) in culture. For instance, the Dicer enzyme may be activated by virtue of being recombinantly expressed or it may be activated by use of an
15   agent which (i) induces expression of the endogenous gene, (ii) stabilizes the protein from degradation, and/or (iii) allosterically modies the enzyme to increase its activity (by altering its Kcat, Km or both).

A. Dicer and Argonaut Activities

20   In certain embodiment, at least one of the activated RNAi enzymes is Dicer, or a homolog thereof. In certain preferred embodiments, the present method provides for ectopic activation of Dicer. As used herein, the term "Dicer" refers to a protein which (a) mediates an RNAi response and (b) has an amino acid sequence at least 50 percent identical, and more preferablty at least 75, 85, 90 or 95 percent identical to SEQ ID No. 2
25   or 4, and/or which can be encoded by a nucleic acid which hybridizes under wash conditions of 2 x SSC at 22°C, and more preferably 0.2 x SSC at 65°C, to a nucleotide represented by SEQ ID No. 1 or 3. Accordingly, the method may comprise introducing a dsRNA contruct into a cell in which Dicer has been recombinantly expressed or otherwise ectopically activated.

30   In certain embodiment, at least one of the activated RNAi enzymes is Argonaut, or a homolog thereof. In certain preferred embodiments, the present method provides for ectopic activation of Argonaut. As used herein, the term "Argonaut" refers to a protein which (a) mediates an RNAi response and (b) has an amino acid sequence at least 50 percent identical, and more preferablty at least 75, 85, 90 or 95 percent identical to the
35   amino acid sequence shown in Figure 24. Accordingly, the method may comprise

introducing a dsRNA contruct into a cell in which Argonaut has been recombinantly expressed or otherwise ectopically activated.

This invention also provides expression vectors containing a nucleic acid encoding a Dicer or Argonaut polypeptides, operably linked to at least one transcriptional regulatory sequence. Operably linked is intended to mean that the nucleotide sequence is linked to a regulatory sequence in a manner which allows expression of the nucleotide sequence. Regulatory sequences are art-recognized and are selected to direct expression of the subject Dicer or Argonaut proteins. Accordingly, the term transcriptional regulatory sequence includes promoters, enhancers and other expression control elements. Such regulatory sequences are described in Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). For instance, any of a wide variety of expression control sequences, sequences that control the expression of a DNA sequence when operatively linked to it, may be used in these vectors to express DNA sequences encoding Dicer or Argonaut polypeptides of this invention. Such useful expression control sequences, include, for example, a viral LTR, such as the LTR of the Moloney murine leukemia virus, the early and late promoters of SV40, adenovirus or cytomegalovirus immediate early promoter, the lac system, the trp system, the TAC or TRC system, T7 promoter whose expression is directed by T7 RNA polymerase, the major operator and promoter regions of phage λ, the control regions for fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase, e.g., Pho5, the promoters of the yeast α-mating factors, the polyhedron promoter of the baculovirus system and other sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof. It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the type of protein desired to be expressed.

Moreover, the vector's copy number, the ability to control that copy number and the expression of any other proteins encoded by the vector, such as antibiotic markers, should also be considered.

The recombinant Dicer or Argonaut genes can be produced by ligating nucleic acid encoding a Dicer or Argonaut polypeptide into a vector suitable for expression in either prokaryotic cells, eukaryotic cells, or both. Expression vectors for production of recombinant forms of the subject Dicer or Argonaut polypeptides include plasmids and other vectors. For instance, suitable vectors for the expression of a Dicer or Argonaut polypeptide include plasmids of the types: pBR322-derived plasmids, pEMBL-derived

plasmids, pEX-derived plasmids, pBTac-derived plasmids and pUC-derived plasmids for expression in prokaryotic cells, such as E. coli.

A number of vectors exist for the expression of recombinant proteins in yeast. For instance, YEP24, YIP5, YEP51, YEP52, pYES2, and YRP17 are cloning and expression vehicles useful in the introduction of genetic constructs into S. cerevisiae (see, for example, Broach et al. (1983) in Experimental Manipulation of Gene Expression, ed. M. Inouye Academic Press, p. 83, incorporated by reference herein). These vectors can replicate in E. coli due the presence of the pBR322 ori, and in S. cerevisiae due to the replication determinant of the yeast 2 micron plasmid. In addition, drug resistance markers such as ampicillin can be used. In an illustrative embodiment, a Dicer or Argonaut polypeptide is produced recombinantly utilizing an expression vector generated by sub-cloning the coding sequence of a Dicer or Argonaut gene.

The preferred mammalian expression vectors contain both prokaryotic sequences, to facilitate the propagation of the vector in bacteria, and one or more eukaryotic transcription units that are expressed in eukaryotic cells. The pcDNAI/amp, pcDNAI/neo, pRc/CMV, pSV2gpt, pSV2neo, pSV2-dhfr, pTk2, pRSVneo, pMSG, pSVT7, pko-neo and pHyg derived vectors are examples of mammalian expression vectors suitable for transfection of eukaryotic cells. Some of these vectors are modified with sequences from bacterial plasmids, such as pBR322, to facilitate replication and drug resistance selection in both prokaryotic and eukaryotic cells. Alternatively, derivatives of viruses such as the bovine papillomavirus (BPV-1), or Epstein-Barr virus (pHEBo, pREP-derived and p205) can be used for transient expression of proteins in eukaryotic cells. The various methods employed in the preparation of the plasmids and transformation of host organisms are well known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells, as well as general recombinant procedures, see Molecular Cloning A Laboratory Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989) Chapters 16 and 17.

In yet another embodiment, the subject invention provides a "gene activation" construct which, by homologous recombination with a genomic DNA, alters the transcriptional regulatory sequences of an endogenous Dicer or Argonaut gene. For instance, the gene activation construct can replace the endogenous promoter of a Dicer or Argonaut gene with a heterologous promoter, e.g., one which causes constitutive expression of the Dicer or Argonaut gene or which causes inducible expression of the gene under conditions different from the normal expression pattern of Dicer or Argonaut. A variety of different formats for the gene activation constructs are available. See, for

example, the Transkaryotic Therapies, Inc PCT publications WO93/09222, WO95/31560, WO96/29411, WO95/31560 and WO94/12650.

In preferred embodiments, the nucleotide sequence used as the gene activation construct can be comprised of (1) DNA from some portion of the endogenous Dicer or
5   Argonaut gene (exon sequence, intron sequence, promoter sequences, etc.) which direct recombination and (2) heterologous transcriptional regulatory sequence(s) which is to be operably linked to the coding sequence for the genomic Dicer or Argonaut gene upon recombination of the gene activation construct. For use in generating cultures of Dicer or Argonaut producing cells, the construct may further include a reporter gene to detect the
10   presence of the knockout construct in the cell.

The gene activation construct is inserted into a cell, and integrates with the genomic DNA of the cell in such a position so as to provide the heterologous regulatory sequences in operative association with the native Dicer or Argonaut gene. Such insertion occurs by homologous recombination, i.e., recombination regions of the activation
15   construct that are homologous to the endogenous Dicer or Argonaut gene sequence hybridize to the genomic DNA and recombine with the genomic sequences so that the construct is incorporated into the corresponding position of the genomic DNA.

The terms "recombination region" or "targeting sequence" refer to a segment (i.e., a portion) of a gene activation construct having a sequence that is substantially identical to
20   or substantially complementary to a genomic gene sequence, e.g., including 5' flanking sequences of the genomic gene, and can facilitate homologous recombination between the genomic sequence and the targeting transgene construct.

As used herein, the term "replacement region" refers to a portion of a activation construct which becomes integrated into an endogenous chromosomal location following
25   homologous recombination between a recombination region and a genomic sequence.

The heterologous regulatory sequences, e.g., which are provided in the replacement region, can include one or more of a variety elements, including: promoters (such as constitutive or inducible promoters), enhancers, negative regulatory elements, locus control regions, transcription factor binding sites, or combinations thereof.
30   Promoters/enhancers which may be used to control the expression of the targeted gene *in vivo* include, but are not limited to, the cytomegalovirus (CMV) promoter/enhancer (Karasuyama et al., 1989, *J. Exp. Med.*, 169:13), the human β-actin promoter (Gunning et al. (1987) *PNAS* 84:4831-4835), the glucocorticoid-inducible promoter present in the mouse mammary tumor virus long terminal repeat (MMTV LTR)
35   (Klessig et al. (1984) *Mol. Cell Biol.* 4:1354-1362), the long terminal repeat sequences of Moloney murine leukemia virus (MuLV LTR) (Weiss et al. (1985) *RNA Tumor Viruses,*

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York), the SV40 early or late region promoter (Bernoist et al. (1981) *Nature* 290:304-310; Templeton et al. (1984) *Mol. Cell Biol.*, 4:817; and Sprague et al. (1983) *J. Virol.*, 45:773), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (RSV) (Yamamoto et al., 1980, *Cell*,

5   22:787-797), the herpes simplex virus (HSV) thymidine kinase promoter/enhancer (Wagner et al. (1981) *PNAS* 82:3567-71), and the herpes simplex virus LAT promoter (Wolfe et al. (1992) *Nature Genetics*, 1:379-384).

In still other embodiments, the replacement region merely deletes a negative transcriptional control element of the native gene, e.g., to activate expression, or ablates a

10   positive control element, e.g., to inhibit expression of the targeted gene.

B. Cell/Organism

The cell with the target gene may be derived from or contained in any organism (e.g., plant, animal, protozoan, virus, bacterium, or fungus). The dsRNA construct may be

15   synthesized either in vivo or in vitro. Endogenous RNA polymerase of the cell may mediate transcription in vivo, or cloned RNA polymerase can be used for transcription in vivo or in vitro. For generating double stranded transcripts from a transgene in vivo, a regulatory region may be used to transcribe the RNA strand (or strands).

Furthermore, genetic manipulation becomes possible in organisms that are not

20   classical genetic models. Breeding and screening programs may be accelerated by the ability to rapidly assay the consequences of a specific, targeted gene disruption. Gene disruptions may be used to discover the function of the target gene, to produce disease models in which the target gene are involved in causing or preventing a pathological condition, and to produce organisms with improved economic properties.

25   The cell with the target gene may be derived from or contained in any organism. The organism may a plant, animal, protozoan, bacterium, virus, or fungus. The plant may be a monocot, dicot or gymnosperm; the animal may be a vertebrate or invertebrate. Preferred microbes are those used in agriculture or by industry, and those that are pathogenic for plants or animals. Fungi include organisms in both the mold and yeast

30   morphologies.

Plants include arabidopsis; field crops (e.g., alfalfa, barley, bean, corn, cotton, flax, pea, rape, rice, rye, safflower, sorghum, soybean, sunflower, tobacco, and wheat); vegetable crops (e.g., asparagus, beet, broccoli, cabbage, carrot, cauliflower, celery, cucumber, eggplant, lettuce, onion, pepper, potato, pumpkin, radish, spinach, squash, taro,

35   tomato, and zucchini); fruit and nut crops (e.g., almond, apple, apricot, banana, blackberry,

blueberry, cacao, cherry, coconut, cranberry, date, faJoa, filbert, grape, grapefruit, guava, kiwi, lemon, lime, mango, melon, nectarine, orange, papaya, passion fruit, peach, peanut, pear, pineapple, pistachio, plum, raspberry, strawberry, tangerine, walnut, and watermelon); and ornamentals (e.g., alder, ash, aspen, azalea, birch, boxwood, camellia, carnation, chrysanthemum, elm, fir, ivy, jasmine, juniper, oak, palm, poplar, pine, redwood, rhododendron, rose, and rubber).

Examples of vertebrate animals include fish, mammal, cattle, goat, pig, sheep, rodent, hamster, mouse, rat, primate, and human.

Invertebrate animals include nematodes, other worms, drosophila, and other insects. Representative generae of nematodes include those that infect animals (e.g., Ancylostoma, Ascaridia, Ascaris, Bunostomum, Caenorhabditis, Capillaria, Chabertia, Cooperia, Dictyocaulus, Haernonchus, Heterakis, Nematodirus, Oesophagostomum, Ostertagia, Oxyuris, Parascaris, Strongylus, Toxascaris, Trichuris, Trichostrongylus, Tflichonema, Toxocara, Uncinaria) and those that infect plants (e.g., B ursaphalenchus, Criconerriella, Diiylenchus, Ditylenchus, Globodera, Helicotylenchus, Heterodera, Longidorus, Melodoigyne, Nacobbus, Paratylenchus, Pratylenchus, Radopholus, Rotelynchus, Tylenchus, and Xiphinerna). Representative orders of insects include Coleoptera, Diptera, Lepidoptera, and Homoptera.

The cell having the target gene may be from the germ line or somatic, totipotent or pluripotent, dividing or non-dividing, parenchyma or epithelium, immortalized or transformed, or the like. The cell may be a stem cell or a differentiated cell. Cell types that are differentiated include adipocytes, fibroblasts, myocytes, cardiomyocytes, endothelium, neurons, glia, blood cells, megakaryocytes, lymphocytes, macrophages, neutrophils, eosinophils, basophils, mast cells, leukocytes, granulocytes, keratinocytes, chondrocytes, osteoblasts, osteoclasts, hepatocytes, and cells of the endocrine or exocrine glands.

C. Targeted Genes

The target gene may be a gene derived from the cell, an endogenous gene, a transgene, or a gene of a pathogen which is present in the cell after infection thereof. Depending on the particular target gene and the dose of double stranded RNA material delivered, the procedure may provide partial or complete loss of function for the target gene. Lower doses of injected material and longer times after administration of dsRNA may result in inhibition in a smaller fraction of cells. Quantitation of gene expression in a cell may show similar amounts of inhibition at the level of accumulation of target mRNA or translation of target protein.

"Inhibition of gene expression" refers to the absence (or observable decrease) in the level of protein and/or mRNA product from a target gene. "Specificity" refers to the ability to inhibit the target gene without manifest effects on other genes of the cell. The consequences of inhibition can be confirmed by examination of the outward properties of

5   the cell or organism (as presented below in the examples) or by biochemical techniques such as RNA solution hybridization, nuclease protection, Northern hybridization, reverse transcription, gene expression monitoring with a microarray, antibody binding, enzyme linked immunosorbent assay (ELISA), Western blotting, radioImmunoassay (RIA), other immunoassays, and fluorescence activated cell analysis (FACS). For RNA-mediated

10   inhibition in a cell line or whole organism, gene expression is conveniently assayed by use of a reporter or drug resistance gene whose protein product is easily assayed. Such reporter genes include acetohydroxyacid synthase (AHAS), alkaline phosphatase (AP), beta galactosidase (LacZ), beta glucoronidase (GUS), chloramphenicol acetyltransferase (CAT), green fluorescent protein (GFP), horseradish peroxidase (HRP), luciferase (Luc),

15   nopaline synthase (NOS), octopine synthase (OCS), and derivatives thereof multiple selectable markers are available that confer resistance to ampicillin, bleomycin, chloramphenicol, gentamycin, hygromycin, kanamycin, lincomycin, methotrexate, phosphinothricin, puromycin, and tetracyclin.

Depending on the assay, quantitation of the amount of gene expression allows one

20   to determine a degree of inhibition which is greater than 10%, 33%, 50%, 90%, 95% or 99% as compared to a cell not treated according to the present invention. Lower doses of injected material and longer times after administration of dsRNA may result in inhibition in a smaller fraction of cells (e.g., at least 10%, 20%, 50%, 75%,90%, or 95% of targeted cells). Quantitation of gene expression in a cell may show similar amounts of inhibition at

25   the level of accumulation of target mRNA or translation of target protein. As an example, the efficiency of inhibition may be determined by assessing the amount of gene product in the cell: mRNA may be detected with a hybridization probe having a nucleotide sequence outside the region used for the inhibitory double-stranded RNA, or translated polypeptide may be detected with an antibody raised against the polypeptide sequence of that region.

30   As disclosed herein, the present invention may is not limited to any type of target gene or nucleotide sequence. But the following classes of possible target genes are listed for illustrative purposes: developmental genes (e.g., adhesion molecules, cyclin kinase inhibitors, Writ family members, Pax family members, Winged helix family members, Hox family members, cytokines/lymphokines and their receptors, growth/differentiation

35   factors and their receptors, neurotransmitters and their receptors); oncogenes (e.g., ABLI, BCLI, BCL2, BCL6, CBFA2, CBL, CSFIR, ERBA, ERBB, EBRB2, ETSI, ETS1, ETV6, FGR, FOS, FYN, HCR, HRAS, JUN, KRAS, LCK, LYN, MDM2, MLL, MYB, MYC,

MYCLI, MYCN, NRAS, PIM 1, PML, RET, SRC, TALI, TCL3, and YES); tumor suppressor genes (e.g., APC, BRCA 1, BRCA2, MADH4, MCC, NF 1, NF2, RB 1, TP53, and WTI); and enzymes (e.g., ACC synthases and oxidases, ACP desaturases and hydroxylases, ADP-glucose pyrophorylases, ATPases, alcohol dehydrogenases, amylases, amyloglucosidases, catalases, cellulases, chalcone synthases, chitinases, cyclooxygenases, decarboxylases, dextrinases, DNA and RNA polymerases, galactosidases, glucanases, glucose oxidases, granule-bound starch synthases, GTPases, helicases, hemicellulases, integrases, inulinases, invertases, isomerases, kinases, lactases, lipases, lipoxygenases, lysozymes, nopaline synthases, octopine synthases, pectinesterases, peroxidases, phosphatases, phospholipases, phosphorylases, phytases, plant growth regulator synthases, polygalacturonases, proteinases and peptidases, pullanases, recombinases, reverse transcriptases, RUBISCOs, topoisomerases, and xylanases).

D. dsRNA constructs

The dsRNA construct may comprise one or more strands of polymerized ribonucleotide. It may include modifications to either the phosphate-sugar backbone or the nucleoside. For example, the phosphodiester linkages of natural RNA may be modified to include at least one of a nitrogen or sulfur heteroatom. Modifications in RNA structure may be tailored to allow specific genetic inhibition while avoiding a general panic response in some organisms which is generated by dsRNA. Likewise, bases may be modified to block the activity of adenosine deaminase. The dsRNA construct may be produced enzymatically or by partial/total organic synthesis, any modified ribonucleotide can be introduced by in vitro enzymatic or organic synthesis.

The dsRNA construct may be directly introduced into the cell (i.e., intracellularly); or introduced extracellularly into a cavity, interstitial space, into the circulation of an organism, introduced orally, or may be introduced by bathing an organism in a solution containing RNA. Methods for oral introduction include direct mixing of RNA with food of the organism, as well as engineered approaches in which a species that is used as food is engineered to express an RNA, then fed to the organism to be affected. Physical methods of introducing nucleic, acids include injection directly into the cell or extracellular injection into the organism of an RNA solution.

The double-stranded structure may be formed by a single self-complementary RNA strand or two complementary RNA strands. RNA duplex formation may be initiated either inside or outside the cell. The RNA may be introduced in an amount which allows delivery of at least one copy per cell. Higher doses (e.g., at least 5, 10, 100, 500 or 1000 copies per cell) of double-stranded material may yield more effective inhibition; lower

doses may also be useful for specific applications. Inhibition is sequence-specific in that nucleotide sequences corresponding to the duplex region of the RNA are targeted for genetic inhibition.

5   dsRNA constructs containing a nucleotide sequences identical to a portion of the target gene are preferred for inhibition. RNA sequences with insertions, deletions, and single point mutations relative to the target sequence have also been found to be effective for inhibition. Thus, sequence identity may optimized by sequence comparison and alignment algorithms known in the art (see Gribskov and Devereux, Sequence Analysis Primer, Stockton Press, 199 1, and references cited therein) and calculating the percent

10   difference between the nucleotide sequences by, for example, the Smith-Waterman algorithm as implemented in the BESTFIT software program using default parameters (e.g., University of Wisconsin Genetic Computing Group). Greater than 90% sequence identity, or even 100% sequence identity, between the inhibitory RNA and the portion of the target gene is preferred. Alternatively, the duplex region of the RNA may be defined

15   functionally as a nucleotide sequence that is capable of hybridizing with a portion of the target gene transcript (e.g., 400 mM NaCl, 40 mM PIPES pH 6.4, 1 mM EDTA, 50'C or 70'C hybridization for 12-16 hours; followed by washing). The length of the identical nucleotide sequences may be, for example, at least 25, 50, 100, 200, 300 or 400 bases. In certain embodiments, the dsRNA construct is 400-800 bases in length.

20   100% sequence identity between the RNA and the target gene is not required to practice the present invention. Thus the invention has the advantage of being able to tolerate sequence variations that might be expected due to genetic mutation, strain polymorphism, or evolutionary divergence.

The dsRNA construct may be synthesized either in vivo or in vitro. Endogenous

25   RNA polyrnerase of the cell may mediate transcription in vivo, or cloned RNA polymerase can be used for transcription in vivo or in vitro. For transcription from a transgene in vivo or an expression construct, a regulatory region (e.g., promoter, enhancer, silencer, splice donor and acceptor, polyadenylation) may be used to transcribe the dsRNA strand (or strands). Inhibition may be targeted by specific transcription in an organ, tissue,

30   or cell type; stimulation of an environmental condition (e.g., infection, stress, temperature, chemical inducers); and/or engineering transcription at a developmental stage or age. The RNA strands may or may not be polyadenylated; the RNA strands may or may not be capable of being translated into a polypeptide by a cell's translational apparatus. The dsRNA construct may be chemically or enzymatically synthesized by manual or

35   automated reactions. The dsRNA construct may be synthesized by a cellular RNA polymerase or a bacteriophage RNA polymerase (e.g., T3, T7, SP6). The use and

production of an expression construct are known in the art32,33,34 (see also WO 97/32016; U.S. Pat. Nos. 5,593,874, 5,698,425, 5,712,135, 5,789,214, and 5,804,693; and the references cited therein). If synthesized chemically or by in vitro enzymatic synthesis, the RNA may be purified prior to introduction into the cell. For example, RNA can be
5   punified from a mixture by extraction with a solvent or resin, precipitation, electrophoresis, chromatography or a combination thereof. Alternatively, the dsRNA construct may be used with no or a minimum of purification to avoid losses due to sample processing.. The dsRNA construct may be dried for storage or dissolved in an aqueous solution. The solution may contain buffers or salts to promote annealing, and/or
10   stabilization of the duplex strands.

Physical methods of introducing nucleic acids include injection of a solution containing the dsRNA construct, bombardment by particles covered by the dsRNA construct, soaking the cell or organism in a solution of the RNA, or electroporation of cell membranes in the presence of the dsRNA construct. A viral construct packaged into a viral
15   particle would accomplish both efficient introduction of an expression construct into the cell and transcription of dsRNA construct encoded by the expression construct. Other methods known in the art for introducing nucleic acids to cells may be used, such as lipid-mediated carrier transport, chemicalmediated transport, such as calcium phosphate, and the like. Thus the dsRNA construct may be introduced along with components that
20   perform one or more of the following activities: enhance RNA uptake by the cell, promote annealing of the duplex strands, stabilize the annealed strands, or other-wise increase inhibition of the target gene.

E. Illustrative Uses

25   One utility of the present invention is as a method of identifying gene function in an organism, especially higher eukaryotes comprising the use of double-stranded RNA to inhibit the activity of a target gene of previously unknown function. Instead of the time consuming and laborious isolation of mutants by traditional genetic screening, functional genomics would envision determining the function of uncharacterized genes by employing
30   the invention to reduce the amount and/or alter the timing of target gene activity. The invention could be used in determining potential targets for pharmaceutics, understanding normal and pathological events associated with development, determining signaling pathways responsible for postnatal development/aging, and the like. The increasing speed of acquiring nucleotide sequence information from genomic and expressed gene sources,
35   including total sequences for mammalian genomes, can be coupled with the invention to determine gene function in a cell or in a whole organism. The preference of different

organisms to use particular codons, searching sequence databases for related gene products, correlating the linkage map of genetic traits with the physical map from which the nucleotide sequences are derived, and artificial intelligence methods may be used to define putative open reading frames from the nucleotide sequences acquired in such sequencing projects.

A simple assay would be to inhibit gene expression according to the partial sequence available from an expressed sequence tag (EST). Functional alterations in growth, development, metabolism, disease resistance, or other biological processes would be indicative of the normal role of the EST's gene product.

The ease with which the dsRNA construct can be introduced into an intact cell/organism containing the target gene allows the present invention to be used in high throughput screening (HTS). For example, duplex RNA can be produced by an amplification reaction using primers flanking the inserts of any gene library derived from the target celUorganism. Inserts may be derived from genomic DNA or mRNA (e.g., cDNA and cRNA). Individual clones from the library can be replicated and then isolated in separate reactions, but preferably the library is maintained in individual reaction vessels (e.g., a 96 well microtiter plate) to minimize the number of steps required to practice the invention and to allow automation of the process. Solutions containing duplex RNAs that are capable of inhibiting the different expressed genes can be placed into individual wells positioned on a microtiter plate as an ordered array, and intact cells/organisms in each well can be assayed for any changes or modifications in behavior or development due to inhibition of target gene activity. The amplified RNA can be fed directly to, injected into, the cell/organism containing the target gene. Alternatively, the duplex RNA can be produced by in vivo or in vitro transcription from an expression construct used to produce the library. The construct can be replicated as individual clones of the library and transcribed to produce the RNA; each clone can then be fed to, or injected into, the cell/organism containing the target gene. The function of the target gene can be assayed from the effects it has on the cell/organism when gene activity is inhibited. This screening could be amenable to small subjects that can be processed in large number, for example, tissue culture cells derived from mammals, especially primates, and most preferably humans.

If a characteristic of an organism is determined to be genetically linked to a polymorphism through RFLP or QTL analysis, the present invention can be used to gain insight regarding whether that genetic polymorphism might be directly responsible for the characteristic. For example, a fragment defining the genetic polymorphism or sequences in the vicinity of such a genetic polymorphism can be amplified to produce an RNA, the

- 24 -

duplex RNA can be introduced to the organism or cell, and whether an alteration in the characteristic is correlated with inhibition can be determined. Of course, there may be trivial explanations for negative results with this type of assay, for example: inhibition of the target gene causes lethality, inhibition of the target gene may not result in any

5   observable alteration, the fragment contains nucleotide sequences that are not capable of inhibiting the target gene, or the target gene's activity is redundant.

The present invention may be useful in allowing the inhibition of essential genes. Such genes may be required for cell or organism viability at only particular stages of development or cellular compartments. The functional equivalent of conditional mutations

10   may be produced by inhibiting activity of the target gene when or where it is not required for viability. The invention allows addition of RNA at specific times of development and locations in the organism without introducing permanent mutations into the target genome.

If alternative splicing produced a family of transcripts that were distinguished by usage of characteristic exons, the present invention can target inhibition through the

15   appropriate exons to specifically inhibit or to distinguish among the functions of family members. For example, a hormone that contained an alternatively spliced transmembrane domain may be expressed in both membrane bound and secreted forms. Instead of isolating a nonsense mutation that terminates translation before the transmembrane domain, the functional consequences of having only secreted hormone can be determined

20   according to the invention by targeting the exon containing the transmembrane domain and thereby inhibiting expression of membrane-bound hormone.

The present invention may be used alone or as a component of a kit having at least one of the reagents necessary to carry out the in vitro or in vivo introduction of RNA to test samples or subjects. Preferred components are the dsRNA and a vehicle that promotes

25   introduction of the dsRNA. Such a kit may also include instructions to allow a user of the kit to practice the invention.

Alternatively, an organism may be engineered to produce dsRNA which produces commercially or medically beneficial results, for example, resistance to a pathogen or its pathogenic effects, improved growth, or novel developmental patterns.

30

IV. *Exemplification*

The invention, now being generally described, will be more readily understood by reference to the following examples, which are included merely for purposes of illustration of certain aspects and embodiments of the present invention and are not intended to limit

35   the invention.

- 25 -

Example 1: An RNA-directed nuclease mediates RNAi gene silencing

In a diverse group of organisms that includes *Caenorhabditis elegans*, *Drosophila*, planaria, hydra, trypanosomes, fungi and plants, the introduction of double-stranded RNAs

5   inhibits gene expression in a sequence-specific manner[1-7]. These responses, called RNA interference or post-transcriptional gene silencing, may provide anti-viral defence, modulate transposition or regulate gene expression[1, 6, 8-10]. We have taken a biochemical approach towards elucidating the mechanisms underlying this genetic phenomenon. Here we show that 'loss-of-function' phenotypes can be created in cultured *Drosophila* cells by

10   transfection with specific double-stranded RNAs. This coincides with a marked reduction in the level of cognate cellular messenger RNAs. Extracts of transfected cells contain a nuclease activity that specifically degrades exogenous transcripts homologous to transfected double-stranded RNA. This enzyme contains an essential RNA component. After partial purification, the sequence-specific nuclease co-fractionates with a discrete,

15   ~25-nucleotide RNA species which may confer specificity to the enzyme through homology to the substrate mRNAs.

Although double-stranded RNAs (dsRNAs) can provoke gene silencing in numerous biological contexts including *Drosophila*[11, 12], the mechanisms underlying this phenomenon have remained mostly unknown. We therefore wanted to establish a

20   biochemically tractable model in which such mechanisms could be investigated.

Transient transfection of cultured, *Drosophila* S2 cells with a *lacZ* expression vector resulted in β-galactosidase activity that was easily detectable by an *in situ* assay (Fig. 1a). This activity was greatly reduced by co-transfection with a dsRNA corresponding to the first 300 nucleotides of the *lacZ* sequence, whereas co-transfection

25   with a control dsRNA (*CD8*) (Fig. 1a) or with single-stranded RNAs of either sense or antisense orientation (data not shown) had little or no effect. This indicated that dsRNAs could interfere, in a sequence-specific fashion, with gene expression in cultured cells.

To determine whether RNA interference (RNAi) could be used to target endogenous genes, we transfected S2 cells with a dsRNA corresponding to the first 540

30   nucleotides of *Drosophila cyclin E*, a gene that is essential for progression into S phase of the cell cycle. During log-phase growth, untreated S2 cells reside primarily in G2/M (Fig. 1b). Transfection with *lacZ* dsRNA had no effect on cell-cycle distribution, but transfection with the *cyclin E* dsRNA caused a G1-phase cell-cycle arrest (Fig. 1b). The ability of *cyclin E* dsRNA to provoke this response was length-dependent. Double-

35   stranded RNAs of 540 and 400 nucleotides were quite effective, whereas dsRNAs of 200 and 300 nucleotides were less potent. Double-stranded *cyclin E* RNAs of 50 or 100

nucleotides were inert in our assay, and transfection with a single-stranded, antisense *cyclin E* RNA had virtually no effect.

One hallmark of RNAi is a reduction in the level of mRNAs that are homologous to the dsRNA. Cells transfected with the *cyclin E* dsRNA (bulk population) showed
5   diminished endogenous *cyclin E* mRNA as compared with control cells (Fig. 1c). Similarly, transfection of cells with dsRNAs homologous to *fizzy*, a component of the anaphase-promoting complex (APC) or *cyclin A*, a cyclin that acts in S, G2 and M, also caused reduction of their cognate mRNAs (Fig. 1c). The modest reduction in *fizzy* mRNA levels in cells transfected with *cyclin A* dsRNA probably resulted from arrest at a point in
10   the division cycle at which *fizzy* transcription is low[14, 15]. These results indicate that RNAi may be a generally applicable method for probing gene function in cultured *Drosophila* cells.

The decrease in mRNA levels observed upon transfection of specific dsRNAs into *Drosophila* cells could be explained by effects at transcriptional or post-transcriptional
15   levels. Data from other systems have indicated that some elements of the dsRNA response may affect mRNA directly (reviewed in refs [1] and [6]). We therefore sought to develop a cell-free assay that reflected, at least in part, RNAi.

S2 cells were transfected with dsRNAs corresponding to either *cyclin E* or *lacZ*. Cellular extracts were incubated with synthetic mRNAs of *lacZ* or *cyclin E*. Extracts
20   prepared from cells transfected with the 540-nucleotide *cyclin E* dsRNA efficiently degraded the *cyclin E* transcript; however, the *lacZ* transcript was stable in these lysates (Fig. 2a). Conversely, lysates from cells transfected with the *lacZ* dsRNA degraded the *lacZ* transcript but left the *cyclin E* mRNA intact. These results indicate that RNAi ablates target mRNAs through the generation of a sequence-specific nuclease activity. We have
25   termed this enzyme RISC (RNA-induced silencing complex). Although we occasionally observed possible intermediates in the degradation process (see Fig. 2), the absence of stable cleavage end-products indicates an exonuclease (perhaps coupled to an endonuclease). However, it is possible that the RNAi nuclease makes an initial endonucleolytic cut and that non-specific exonucleases in the extract complete the
30   degradation process[16]. In addition, our ability to create an extract that targets *lacZ in vitro* indicates that the presence of an endogenous gene is not required for the RNAi response.

To examine the substrate requirements for the dsRNA-induced, sequence-specific nuclease activity, we incubated a variety of *cyclin-E*-derived transcripts with an extract derived from cells that had been transfected with the 540-nucleotide *cyclin E* dsRNA (Fig.
35   2b, c). Just as a length requirement was observed for the transfected dsRNA, the RNAi nuclease activity showed a dependence on the size of the RNA substrate. Both a 600-

nucleotide transcript that extends slightly beyond the targeted region (Fig. 2b) and an ~1-
kilobase (kb) transcript that contains the entire coding sequence (data not shown) were
completely destroyed by the extract. Surprisingly, shorter substrates were not degraded as
efficiently. Reduced activity was observed against either a 300- or a 220-nucleotide

5   transcript, and a 100-nucleotide transcript was resistant to nuclease in our assay. This was
not due solely to position effects because ~100-nucleotide transcripts derived from other
portions of the transfected dsRNA behaved similarly (data not shown). As expected, the
nuclease activity (or activities) present in the extract could also recognize the antisense
strand of the *cyclin E* mRNA. Again, substrates that contained a substantial portion of the

10   targeted region were degraded efficiently whereas those that contained a shorter stretch of
homologous sequence (~130 nucleotides) were recognized inefficiently (Fig. 2c, as600).
For both the sense and antisense strands, transcripts that had no homology with the
transfected dsRNA (Fig. 2b, Eout; Fig. 2c, as300) were not degraded. Although we cannot
exclude the possibility that nuclease specificity could have migrated beyond the targeted

15   region, the resistance of transcripts that do not contain homology to the dsRNA is
consistent with data from *C. elegans*. Double-stranded RNAs homologous to an upstream
cistron have little or no effect on a linked downstream cistron, despite the fact that
unprocessed, polycistronic mRNAs can be readily detected[17, 18]. Furthermore, the nuclease
was inactive against a dsRNA identical to that used to provoke the RNAi response *in vivo*

20   (Fig. 2b). In the *in vitro* system, neither a 5' cap nor a poly(A) tail was required, as such
transcripts were degraded as efficiently as uncapped and non-polyadenylated RNAs.

Gene silencing provoked by dsRNA is sequence specific. A plausible mechanism
for determining specificity would be incorporation of nucleic-acid guide sequences into
the complexes that accomplish silencing[19]. In accord with this idea, pre-treatment of

25   extracts with a Ca^{2+}-dependent nuclease (micrococcal nuclease) abolished the ability of
these extracts to degrade cognate mRNAs (Fig. 3). Activity could not be rescued by
addition of non-specific RNAs such as yeast transfer RNA. Although micrococcal
nuclease can degrade both DNA and RNA, treatment of the extract with DNAse I had no
effect (Fig. 3). Sequence-specific nuclease activity, however, did require protein (data not

30   shown). Together, our results support the possibility that the RNAi nuclease is a
ribonucleoprotein, requiring both RNA and protein components. Biochemical
fractionation (see below) is consistent with these components being associated in extract
rather than being assembled on the target mRNA after its addition.

In plants, the phenomenon of co-suppression has been associated with the

35   existence of small (~25-nucleotide) RNAs that correspond to the gene that is being
silenced[19]. To address the possibility that a similar RNA might exist in *Drosophila* and
guide the sequence-specific nuclease in the choice of substrate, we partially purified our

activity through several fractionation steps. Crude extracts contained both sequence-specific nuclease activity and abundant, heterogeneous RNAs homologous to the transfected dsRNA (Figs 2 and 4a). The RNAi nuclease fractionated with ribosomes in a high-speed centrifugation step. Activity could be extracted by treatment with high salt, and

5   ribosomes could be removed by an additional centrifugation step. Chromatography of soluble nuclease over an anion-exchange column resulted in a discrete peak of activity (Fig. 4b, cyclin E). This retained specificity as it was inactive against a heterologous mRNA (Fig. 4b, lacZ). Active fractions also contained an RNA species of 25 nucleotides that is homologous to the cyclin E target (Fig. 4b, northern). The band observed on

10   northern blots may represent a family of discrete RNAs because it could be detected with probes specific for both the sense and antisense cyclin E sequences and with probes derived from distinct segments of the dsRNA (data not shown). At present, we cannot determine whether the 25-nucleotide RNA is present in the nuclease complex in a double-stranded or single-stranded form.

15   RNA interference allows an adaptive defence against both exogenous and endogenous dsRNAs, providing something akin to a dsRNA immune response. Our data, and that of others[19], is consistent with a model in which dsRNAs present in a cell are converted, either through processing or replication, into small specificity determinants of discrete size in a manner analogous to antigen processing. Our results suggest that the

20   post-transcriptional component of dsRNA-dependent gene silencing is accomplished by a sequence-specific nuclease that incorporates these small RNAs as guides that target specific messages based upon sequence recognition. The identical size of putative specificity determinants in plants[19] and animals predicts a conservation of both the mechanisms and the components of dsRNA-induced, post-transcriptional gene silencing in

25   diverse organisms. In plants, dsRNAs provoke not only post-transcriptional gene silencing but also chromatin remodelling and transcriptional repression[20, 21]. It is now critical to determine whether conservation of gene-silencing mechanisms also exists at the transcriptional level and whether chromatin remodelling can be directed in a sequence-specific fashion by these same dsRNA-derived guide sequences.

30

Methods
Cell culture and RNA methods S2 (ref. 22) cells were cultured at 27 °C in 90% Schneider's insect media (Sigma), 10% heat inactivated fetal bovine serum (FBS). Cells were transfected with dsRNA and plasmid DNA by calcium phosphate co-precipitation[23].

35   Identical results were observed when cells were transfected using lipid reagents (for example, Superfect, Qiagen). For FACS analysis, cells were additionally transfected with

a vector that directs expression of a green fluorescent protein (GFP)–US9 fusion protein[13]. These cells were fixed in 90% ice-cold ethanol and stained with propidium iodide at 25 μg ml^{-1}. FACS was performed on an Elite flow cytometer (Coulter). For northern blotting, equal loading was ensured by over-probing blots with a control complementary DNA
5   (RP49). For the production of dsRNA, transcription templates were generated by polymerase chain reaction such that they contained T7 promoter sequences on each end of the template. RNA was prepared using the RiboMax kit (Promega). Confirmation that RNAs were double stranded came from their complete sensitivity to RNAse III (a gift from A. Nicholson). Target mRNA transcripts were synthesized using the Riboprobe kit
10   (Promega) and were gel purified before use.

Extract preparation Log-phase S2 cells were plated on 15-cm tissue culture dishes and transfected with 30 μg dsRNA and 30 μg carrier plasmid DNA. Seventy-two hours after transfection, cells were harvested in PBS containing 5 mM EGTA washed twice in PBS and once in hypotonic buffer (10 mM HEPES pH 7.3, 6 mM β-mercaptoethanol). Cells
15   were suspended in 0.7 packed-cell volumes of hypotonic buffer containing *Complete* protease inhibitors (Boehringer) and 0.5 units ml^{-1} of RNasin (Promega). Cells were disrupted in a dounce homogenizer with a type B pestle, and lysates were centrifuged at 30,000g for 20 min. Supernatants were used in an *in vitro* assay containing 20 mM HEPES pH 7.3, 110 mM KOAc, 1 mM Mg(OAc)$_2$, 3 mM EGTA, 2 mM CaCl$_2$, 1 mM DTT.
20   Typically, 5 μl extract was used in a 10 μl assay that contained also 10,000 c.p.m. synthetic mRNA substrate.

Extract fractionation Extracts were centrifuged at 200,000g for 3 h and the resulting pellet (containing ribosomes) was extracted in hypotonic buffer containing also 1 mM MgCl$_2$ and 300 mM KOAc. The extracted material was spun at 100,000g for 1 h and the
25   resulting supernatant was fractionated on Source 15Q column (Pharmacia) using a KCl gradient in buffer A (20 mM HEPES pH 7.0, 1 mM dithiothreitol, 1 mM MgCl$_2$). Fractions were assayed for nuclease activity as described above. For northern blotting, fractions were proteinase K/SDS treated, phenol extracted, and resolved on 15% acrylamide 8M urea gels. RNA was electroblotted onto Hybond N+ and probed with
30   strand-specific riboprobes derived from cyclin E mRNA. Hybridization was carried out in 500 mM NaPO$_4$ pH 7.0, 15% formamide, 7% SDS, 1% BSA. Blots were washed in 1 SSC at 37–45 °C.

References cited in Example 1

35   1.   Sharp, P. A. RNAi and double-strand RNA. *Genes Dev.* **13**, 139-141 (1999).

2. Sanchez-Alvarado, A. & Newmark, P. A. Double-stranded RNA specifically disrupts gene expression during planarian regeneration. *Proc. Natl Acad. Sci. USA* **96**, 5049-5054 (1999).

3. Lohmann, J. U., Endl, I. & Bosch, T. C. Silencing of developmental genes in Hydra. *Dev. Biol.* **214**, 211-214 (1999).

4. Cogoni, C. & Macino, G. Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase. *Nature* **399**, 166-169 (1999).

5. Waterhouse, P. M., Graham, M. W. & Wang, M. B. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc. Natl Acad. Sci. USA* **95**, 13959-13964 (1998).

6. Montgomery, M. K. & Fire, A. Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression. *Trends Genet.* **14**, 225-228 (1998).

7. Ngo, H., Tschudi, C., Gull, K. & Ullu, E. Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. *Proc. Natl Acad. Sci. USA* **95**, 14687-14692 (1998).

8. Tabara, H. et al. The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell* **99**, 123-132 (1999).

9. Ketting, R. F., Haverkamp, T. H. A., van Luenen, H. G. A. M. & Plasterk, R. H. A. mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner Syndrome helicase and RnaseD. *Cell* **99**, 133-141 (1999).

10. Ratcliff, F., Harrison, B. D. & Baulcombe, D. C. A similarity between viral defense and gene silencing in plants. *Science* **276**, 1558-1560 (1997).

11. Kennerdell, J. R. & Carthew, R. W. Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. *Cell* **95**, 1017-1026 (1998).

12. Misquitta, L. & Paterson, B. M. Targeted disruption of gene function in Drosophila by RNA interference: a role for nautilus in embryonic somatic muscle formation. *Proc. Natl Acad. Sci. USA* **96**, 1451-1456 (1999).

13. Kalejta, R. F., Brideau, A. D., Banfield, B. W. & Beavis, A. J. An integral membrane green fluorescent protein marker, Us9-GFP, is quantitatively retained in cells during propidium iodine-based cell cycle analysis by flow cytometry. *Exp. Cell. Res.* **248**, 322-328 (1999).

14. Wolf, D. A. & Jackson, P. K. Cell cycle: oiling the gears of anaphase. *Curr. Biol.* **8**, R637-R639 (1998).

15. Kramer, E. R., Gieffers, C., Holz, G., Hengstschlager, M. & Peters, J. M. Activation of the human anaphase-promoting complex by proteins of the CDC20/fizzy family. *Curr. Biol.* **8**, 1207-1210 (1998).

16. Shuttleworth, J. & Colman, A. Antisense oligonucleotide-directed cleavage of mRNA in Xenopus oocytes and eggs. *EMBO J.* **7**, 427-434 (1988).

17. Tabara, H., Grishok, A. & Mello, C. C. RNAi in C. elegans: soaking in the genome sequence. *Science* **282**, 430-432 (1998).

18. Bosher, J. M., Dufourcq, P., Sookhareea, S. & Labouesse, M. RNA interference can target pre-mRNA. Consequences for gene expression in a Caenorhabditis elegans operon. *Genetics* **153**, 1245-1256 (1999).

19. Hamilton, J. A. & Baulcombe, D. C. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**, 950-952 (1999).

20. Jones, L. A., Thomas, C. L. & Maule, A. J. De novo methylation and co-suppression induced by a cytoplasmically replicating plant RNA virus. *EMBO J.* **17**, 6385-6393 (1998).

21. Jones, L. A. et al. RNA-DNA interactions and DNA methylation in post-transcriptional gene silencing. *Plant Cell* **11**, 2291-2301 (1999).

22. Schneider, I. Cell lines derived from late embryonic stages of Drosophila melanogaster. *J. Embryol. Exp. Morpho.* **27**, 353-365 (1972).

23. Di Nocera, P. P. & Dawid, I. B. Transient expression of genes introduced into cultured cells of Drosophila. *Proc. Natl Acad. Sci. USA* **80**, 7095-7098 (1983).

Example 2: Role for a bidentate ribonuclease in the initiation step of RNA interference

Genetic approaches in worms, fungi and plants have identified a group of proteins that are essential for double-stranded RNA-induced gene silencing. Among these are ARGONAUTE family members (e.g. RDE1, QDE2)[9,10,30], recQ-family helicases (MUT-7, QDE3)[11,12], and RNA-dependent RNA polymerases (e.g. EGO-1, QDE1, SGS2/SDE1)[13-16]. While potential roles have been proposed, none of these genes has been assigned a definitive function in the silencing process. Biochemical studies have suggested that PTGS is accomplished by a multicomponent nuclease that targets mRNAs for degradation[6,8,17]. We have shown that the specificity of this complex may derive from the incorporation of a small guide sequence that is homologous to the mRNA substrate[6]. Originally identified in plants that were actively silencing transgenes[7], these ~22 nt. RNAs

have been produced during RNAi *in vitro* using an extract prepared from *Drosophila* embryos[8.] Putative guide RNAs can also be produced in extracts from *Drosophila* S2 cells (Fig. 5a). With the goal of understanding the mechanism of post-transcriptional gene silencing, we have undertaken both biochemical fractionation and candidate gene 5   approaches to identify the enzymes that execute each step of RNAi.

Our previous studies resulted in the partial purification of a nuclease, RISC, that is an effector of RNA interference. See Example 1. This enzyme was isolated from *Drosophila* S2 cells in which RNAi had been initiated *in vivo* by transfection with dsRNA. We first sought to determine whether the RISC enzyme and the enzyme that initiates 10   RNAi via processing of dsRNA into 22mers are distinct activities. RISC activity could be largely cleared from extracts by high-speed centrifugation (100,000xg for 60 min.) while the activity that produces 22mers remained in the supernatant (Fig. 5b,c). This simple fractionation indicated that RISC and the 22mer-generating activity are separable and thus distinct enzymes. However, it seems likely that they might interact at some point during 15   the silencing process.

RNAse III family members are among the few nucleases that show specificity for double-stranded RNA[18]. Analysis of the *Drosophila* and *C. elegans* genomes reveals several types of RNAse III enzymes. First is the canonical RNAse III which contains a single RNAse III signature motif and a double-stranded RNA binding domain (dsRBD; 20   e.g. RNC_CAEEL). Second is a class represented by Drosha[19], a *Drosophila* enzyme that contains two RNAse III motifs and a dsRBD (CeDrosha in *C. elegans*). A third class contains two RNAse III signatures and an amino terminal helicase domain (e.g. *Drosophila* CG4792, CG6493, *C. elegans* K12H4.8), and these had previously been proposed by Bass as candidate RNAi nucleases[20]. Representatives of all three classes 25   were tested for the ability to produce discrete, ~22 nt. RNAs from dsRNA substrates.

Partial digestion of a 500 nt. cyclin E dsRNA with purified, bacterial RNAse III produced a smear of products while nearly complete digestion produced a heterogeneous group of ~11-17 nucleotide RNAs (not shown). In order to test the dual-RNAse III enzymes, we prepared T7 epitope-tagged versions of Drosha and CG4792. These were 30   expressed in transfected S2 cells and isolated by immunoprecipitation using antibody-agarose conjugates. Treatment of the dsRNA with the CG4792 immunoprecipitate yielded ~22 nt. fragments similar to those produced in either S2 or embryo extracts (Fig. 6a). Neither activity in extract nor activity in immunoprecipitates depended on the sequence of the RNA substrate since dsRNAs derived from several genes were processed equivalently 35   (see Supplement 1). Negative results were obtained with Drosha and with immunoprecipitates of a DExH box helicase (Homeless[21]; see Fig 6a,b). Western blotting

confirmed that each of the tagged proteins was expressed and immunoprecipitated similarly (see Supplement 2). Thus, we conclude that CG4792 may carry out the initiation step of RNA interference by producing ~22 nt. guide sequences from dsRNAs. Because of its ability to digest dsRNA into uniformly sized, small RNAs, we have named this

5   enzyme Dicer (*Dcr*). *Dicer* mRNA is expressed in embryos, in S2 cells, and in adult flies, consistent with the presence of functional RNAi machinery in all of these contexts (see Supplement 3).

The possibility that Dicer might be the nuclease responsible for the production of guide RNAs from dsRNAs prompted us to raise an antiserum directed against the carboxy-

10   terminus of the Dicer protein (Dicer-1, CG4792). This antiserum could immunoprecipitate a nuclease activity from either Drosophila embryo extracts or from S2 cell lysates that produced ~22 nt. RNAs from dsRNA substrates (Fig. 6C). The putative guide RNAs that are produced by the Dicer-1 enzyme precisely comigrate with 22mers that are produced in extract and with 22mers that are associated with the RISC enzyme

15   (Fig. 6 D,F). It had previously been shown that the enzyme that produced guide RNAs in Drosophila embryo extracts was ATP-dependent[8]. Depletion of this cofactor resulted in an ~6-fold lower rate of dsRNA cleavage and in the production of RNAs with a slightly lower mobility. Of interest was the fact that both Dicer-1 immunoprecipitates and extracts from S2 cells require ATP for the production of ~22mers (Fig. 6D). We do not observe

20   the accumulation of lower mobility products in these cases, although we do routinely observe these in ATP-depleted embryo extracts. The requirement of this nuclease for ATP is a quite unusual property. We hypothesize that this requirement could indicate that the enzyme may act processively on the dsRNA, with the helicase domain harnessing the energy of ATP hydrolysis both for unwinding guide RNAs and for translocation along the

25   substrate.

Efficient induction of RNA interference in *C. elegans* and in *Drosophila* has several requirements. For example, the initiating RNA must be double-stranded, and it must be several hundred nucleotides in length. To determine whether these requirements are dictated by Dicer, we characterized the ability of extracts and of immunoprecipitated

30   enzyme to digest various RNA substrates. Dicer was inactive against single stranded RNAs regardless of length (see Supplement 4). The enyzme could digest both 200 and 500 nucleotide dsRNAs but was significantly less active with shorter substrates (see Supplement 4). Double-stranded RNAs as short as 35 nucleotides could be cut by the enzyme, albeit very inefficiently (data not shown). In contrast, E. coli RNAse III could

35   digest to completion dsRNAs of 35 or 22 nucleotides (not shown). This suggests that the substrate preferences of the Dicer enzyme may contribute to but not wholly determine the size dependence of RNAi.

To determine whether the Dicer enzyme indeed played a role in RNAi *in vivo*, we sought to deplete Dicer activity from S2 cells and test the effect on dsRNA-induced gene silencing. Transfection of S2 cells with a mixture of dsRNAs homologous to the two Drosophila Dicer genes (CG4792 and CG6493) resulted in an ~6-7 fold reduction of Dicer

5   activity either in whole cell lysates or in Dicer-1 immunoprecipitates (Fig. 7A,B). Transfection with a control dsRNA (murine caspase 9) had no effect. Qualitatively similar results were seen if Dicer was examined by Northern blotting (not shown). Depletion of Dicer in this manner substantially compromised the ability of cells to silence subsequently an exogenous, GFP transgene by RNAi (Fig. 7C). These results indicate that Dicer is

10   involved in RNAi *in vivo*. The lack of complete inhibition of silencing could result from an incomplete suppression of Dicer (which is itself required for RNAi) or could indicate that *in vivo*, guide RNAs can be produced by more than one mechanism (e.g. through the action of RNA-dependent RNA polymerases).

Our results indicate that the process of RNA interference can be divided into at

15   least two distinct steps. According to this model, initiation of PTGS would occur upon processing of a double-stranded RNA by Dicer into ~22 nucleotide guide sequences, although we cannot formally exclude the possibility that another, Dicer-associated nuclease may participate in this process. These guide RNAs would be incorporated into a distinct nuclease complex (RISC) that targets single-stranded mRNAs for degradation. An

20   implication of this model is that guide sequences are themselves derived directly from the dsRNA that triggers the response. In accord with this model, we have demonstrated that ^{32}P-labeled, exogenous dsRNAs that have been introduced into S2 cells by transfection are incorporated into the RISC enzyme as 22 mers (Fig. 7E). However, we cannot exclude the possibility that RNA-dependent RNA polymerases might amplify 22mers once they have

25   been generated or provide an alternative method for producing guide RNAs.

The structure of the Dicer enzyme provokes speculation on the mechanism by which the enzyme might produce discretely sized fragments irrespective of the sequence of the dsRNA (see Supplement 1, Fig. 8a). It has been established that bacterial RNAse III acts on its substrate as a dimer[18,22,23]. Similarly, a dimer of Dicer enzymes may be

30   required for cleavage of dsRNAs into ~22 nt. pieces. According to one model, the cleavage interval would be determined by the physical arrangement of the two RNAse III domains within Dicer enzyme (Fig. 8a). A plausible alternative model would dictate that cleavage was directed at a single position by the two RIII domains in a single Dicer protein. The 22 nucleotide interval could be dictated by interaction of neighboring Dicer

35   enzymes or by translocation along the mRNA substrate. The presence of an integral helicase domain suggests that the products of Dicer cleavage might be single-stranded 22 mers that are incorporated into the RISC enzyme as such.

A notable feature of the Dicer family is its evolutionary conservation. Homologs are found in *C. elegans* (K12H4.8), *Arabidopsis* (e.g., CARPEL FACTORY[24], T25K16.4, AC012328_1), mammals (Helicase-MOI[25]) and *S. pombe* (YC9A_SCHPO) (Fig 8b, see Supplements 6,7 for sequence comparisons). In fact, the human Dicer family member is capable of generating ~22 nt. RNAs from dsRNA substrates (Supplement 5) suggesting that these structurally similar proteins may all share similar biochemical functions. It has been demonstrated that exogenous dsRNAs can affect gene function in early mouse embryos[29], and our results suggest that this regulation may be accomplished by an evolutionarily conserved RNAi machinery.

In addition to RNAseIII and helicase motifs, searches of the PFAM database indicate that each Dicer family member also contains a ZAP domain (Fig 8c)[27]. This sequence was defined based solely upon its conservation in the Zwille/ARGONAUTE/Piwi family that has been implicated in RNAi by mutations *in C. elegans* (Rde-1)[9] and *Neurospora* (Qde-2)[10]. Although the function of this domain is unknown, it is intriguing that this region of homology is restricted to two gene families that participate in dsRNA-dependent silencing. Both the ARGONAUTE and Dicer families have also been implicated in common biological processes, namely the determination of stem-cell fates. A hypomorphic allele of *carpel factory*, a member of the Dicer family in *Arabidopsis*, is characterized by increased proliferation in floral meristems[24]. This phenotype and a number of other characteristic features are also shared by *Arabidopsis ARGONAUTE* (*ago1-1*) mutants[26] (C. Kidner and R. Martiennsen, pers. comm.). These genetic analyses begin to provide evidence that RNAi may be more than a defensive response to unusual RNAs but may also play important roles in the regulation of endogenous genes.

With the identification of Dicer as a catalyst of the initiation step of RNAi, we have begun to unravel the biochemical basis of this unusual mechanism of gene regulation. It will be of critical importance to determine whether the conserved family members from other organisms, particularly mammals, also play a role in dsRNA-mediated gene regulation.

Methods

Plasmid constructs. A full-length cDNA encoding Drosha was obtained by PCR from an EST sequenced by the Berkeley Drosophila genome project. The *Homeless* clone was a gift from Gillespie and Berg (Univ. Washington). The T7 epitope-tag was added to the amino terminus of each by PCR, and the tagged cDNAs were cloned into pRIP, a retroviral vector designed specifically for expression in insect cells (E. Bernstein,

unpublished). In this vector, expression is driven by the *Orgyia pseudotsugata* IE2 promoter (Invitrogen). Since no cDNA was available for CG4792/Dicer, a genomic clone was amplified from a bacmid (BACR23F10; obtained from the BACPAC Resource Center in the Dept. of Human Genetics at the Roswell Park Cancer Institute). Again, during amplification, a T7 epitope tag was added at the amino terminus of the coding sequence. The human Dicer gene was isolated from a cDNA library prepared from HaCaT cells (GJH, unpublished). A T7-tagged version of the complete coding sequence was cloned into pCDNA3 (Invitrogen) for expression in human cells (LinX-A).

Cell culture and extract preparation. *S2 and embryo culture.* S2 cells were cultured at 27°C in 5% CO_2 in Schneider's insect media supplemented with 10% heat inactivated fetal bovine serum (Gemini) and 1% antibiotic-antimycotic solution (Gibco BRL). Cells were harvested for extract preparation at 10×10^6 cells/ml. The cells were washed 1X in PBS and were resuspended in a hypotonic buffer (10 mM Hepes pH 7.0, 2mM MgCl2, 6 mM βME) and dounced. Cell lysates were spun 20,000xg for 20 minutes. Extracts were stored at –80°C. *Drosophila* embryos were reared in fly cages by standard methodologies and were collected every 12 hours. The embryos were dechorionated in 50% chlorox bleach and washed thoroughly with distilled water. Lysis buffer (10mM Hepes, 10mM KCl, 1.5 mM $MgCl_2$, 0.5mM EGTA, 10mM β-glycerophosphate, 1mM DTT, 0.2 mM PMSF) was added to the embryos, and extracts were prepared by homogenization in a tissue grinder. Lysates were spun for two hours at 200,000xg and were frozen at –80°C. LinX-A cells, a highly-transfectable derivative of human 293 cells, (Lin Xie and GJH, unpublished) were maintained in DMEM/10%FCS.

Transfections and immunoprecipitations. S2 cells were transfected using a calcium phosphate procedure essentially as previously described[6]. Transfection rates were ~90% as monitored in controls using an *in situ* β-galactosidase assay. LinX-A cells were also transfected by calcium phosphate co-precipitation. For immunoprecipitations, cells (~ 5×10^6 per IP) were transfected with various clones and lysed three days later in IP buffer (125mM KOAc, 1mM MgOAc, 1mM $CaCl_2$, 5mM EGTA, 20mM Hepes pH 7.0, 1mM DTT, 1% NP-40 plus Complete protease inhibitors (Roche)). Lysates were spun for 10 minutes at 14,000xg and supernatants were added to T7 antibody-agarose beads (Novagen). Antibody binding proceeded for 4 hours at 4°C. Beads were centrifuged and washed in lysis buffer three times, and once in reaction buffer. The Dicer antiserum was raised in rabbits using a KLH-conjugated peptide corresponding to the C-terminal 8 amino acids of Drosophila Dicer-1 (CG4792).

Cleavage reactions. *RNA preparation.* Templates to be transcribed into dsRNA were generated by PCR with forward and reverse primers, each containing a T7 promoter

sequence. RNAs were produced using Riboprobe (Promega) kits and were uniformly labeling during the transcription reaction with ^{32}P-UTP. Single-stranded RNAs were purified from 1% agarose gels. *dsRNA cleavage*. Five microliters of embryo or S2 extracts were incubated for one hour at 30°C with dsRNA in a reaction containing 20mM

5   Hepes pH 7.0, 2mM MgOAc, 2mM DTT, 1mM ATP and 5% Superasin (Ambion). Immunoprecipitates were treated similarly except that a minimal volume of reaction buffer (including ATP and Superasin) and dsRNA were added to beads that had been washed in reaction buffer (see above). For ATP depletion, *Drosophila* embryo extracts were incubated for 20 minutes at 30°C with 2mM glucose and 0.375 U of hexokinase (Roche)

10   prior to the addition of dsRNA.

Northern and Western analysis. Total RNA was prepared from *Drosophila* embryos (0-12 hour), from adult flies, and from S2 cells using Trizol (Lifetech). Messenger RNA was isolated by affinity selection using magnetic oligo-dT beads (Dynal). RNAs were electrophoresed on denaturing formaldehyde/agarose gels, blotted and probed

15   with randomly primed DNAs corresponding to Dicer. For Western analysis, T7-tagged proteins were immunoprecipitated from whole cell lysates in IP buffer using anti-T7-antibody-agarose conjugates. Proteins were released from the beads by boiling in Laemmli buffer and were separated by electrophoresis on 8% SDS PAGE. Following transfer to nitrocellulose, proteins were visualized using an HRP-conjugated anti-T7

20   antibody (Novagen) and chemiluminescent detection (Supersignal, Pierce).

RNAi of Dicer. Drosophila S2 cells were transfected either with a dsRNA corresponding to mouse caspase 9 or with a mixture of two dsRNAs corresponding to Drosophila Dicer-1 and Dicer-2 (CG4792 and CG6493). Two days after the initial transfection, cells were again transfected with a mixture containing a GFP expression

25   plasmid and either luciferase dsRNA or GFP dsRNA as previously described[6]. Cells were assayed for Dicer activity or fluorescence three days after the second transfection. Quantification of fluorescent cells was done on a Coulter EPICS cell sorter after fixation. Control transfections indicated that Dicer activity was not affected by the introduction of caspase 9 dsRNA.

30

References cited Example 2

1.   Baulcombe, D. C. RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. *Plant Mol Biol* 32, 79-88 (1996).

2. Wassenegger, M. & Pelissier, T. A model for RNA-mediated gene silencing in higher

35   plants. *Plant Mol Biol* 37, 349-62 (1998).

3. Montgomery, M. K. & Fire, A. Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression [see comments]. *Trends Genet* 14, 255-8 (1998).

4. Sharp, P. A. RNAi and double-strand RNA. *Genes Dev* 13, 139-41 (1999).

5. Sijen, T. & Kooter, J. M. Post-transcriptional gene-silencing: RNAs on the attack or on the defense? [In Process Citation]. *Bioessays* 22, 520-31 (2000).

6. Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* 404, 293-6 (2000).

7. Hamilton, A. J. & Baulcombe, D. C. A species of small antisense RNA in posttranscriptional gene silencing in plants [see comments]. *Science* 286, 950-2 (1999).

8. Zamore, P. D., Tuschl, T., Sharp, P. A. & Bartel, D. P. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101, 25-33 (2000).

9. Tabara, H. *et al.* The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell* 99, 123-32 (1999).

10. Catalanotto, C., Azzalin, G., Macino, G. & Cogoni, C. Gene silencing in worms and fungi. *Nature* 404, 245 (2000).

11. Ketting, R. F., Haverkamp, T. H., van Luenen, H. G. & Plasterk, R. H. Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* 99, 133-41 (1999).

12. Cogoni, C. & Macino, G. Posttranscriptional gene silencing in Neurospora by a RecQ DNA helicase. *Science* 286, 2342-4 (1999).

13. Cogoni, C. & Macino, G. Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase. *Nature* 399, 166-9 (1999).

14. Smardon, A. *et al.* EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans [published erratum appears in Curr Biol 2000 May 18;10(10):R393-4]. *Curr Biol* 10, 169-78 (2000).

15. Mourrain, P. *et al.* Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* 101, 533-42 (2000).

16. Dalmay, T., Hamilton, A., Rudd, S., Angell, S. & Baulcombe, D. C. An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* 101, 543-53 (2000).

17. Tuschl, T., Zamore, P. D., Lehmann, R., Bartel, D. P. & Sharp, P. A. Targeted mRNA degradation by double-stranded RNA in vitro. *Genes Dev* 13, 3191-7 (1999).

18. Nicholson, A. W. Function, mechanism and regulation of bacterial ribonucleases. *FEMS Microbiol Rev* 23, 371-90 (1999).

19. Filippov, V., Solovyev, V., Filippova, M. & Gill, S. S. A novel type of RNase III family proteins in eukaryotes. *Gene* 245, 213-21 (2000).

20. Bass, B. L. Double-stranded RNA as a template for gene silencing. *Cell* 101, 235-8 (2000).

21. Gillespie, D. E. & Berg, C. A. Homeless is required for RNA localization in Drosophila oogenesis and encodes a new member of the DE-H family of RNA-dependent ATPases. *Genes Dev* 9, 2495-508 (1995).

22. Robertson, H. D., Webster, R. E. & Zinder, N. D. Purification and properties of ribonuclease III from Escherichia coli. *J Biol Chem* 243, 82-91 (1968).

23. Dunn, J. J. RNase III cleavage of single-stranded RNA. Effect of ionic strength on the fideltiy of cleavage. *J Biol Chem* 251, 3807-14 (1976).

24. Jacobsen, S. E., Running, M. P. & Meyerowitz, E. M. Disruption of an RNA helicase/RNAse III gene in Arabidopsis causes unregulated cell division in floral meristems. *Development* 126, 5231-43 (1999).

25. Matsuda, S. *et al.* Molecular cloning and characterization of a novel human gene (HERNA) which encodes a putative RNA-helicase. *Biochim Biophys Acta* 1490, 163-9 (2000).

26. Bohmert, K. *et al.* AGO1 defines a novel locus of Arabidopsis controlling leaf development. *Embo J* 17, 170-80 (1998).

27. Sonnhammer, E. L., Eddy, S. R. & Durbin, R. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* 28, 405-20 (1997).

28. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389-402 (1997).

29. Wianny, F. and Zernicka-Goetz, M. Specific interference with gene function by double-stranded RNA in early mouse development. *Nature Cell Biol.* 2, 70-75 (2000).

30. Fagard, M., Boutet, S., Morel, J.-B., Bellini, C. and Vaucheret, H. Ago-1, Qde-2 and Rde-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. *Proc. Natl. Acad. Sci. USA* 97, 11650-11654 (2000).

Example 3: A simplified method for the creation of hairpin constructs for RNA interference.

In numerous model organisms, double stranded RNAs have been shown to cause
effective and specific suppression of gene function (ref. 1). This response, termed RNA
interference or post-transcriptional gene silencing, has evolved into a highly effective
reverse genetic tool in *C. elegans*, *Drosophila*, plants and numerous other systems. In
these cases, double-stranded RNAs can be introduced by injection, transfection or feeding;
however, in all cases, the response is both transient and systemic. Recently, stable
interference with gene expression has been achieved by expression of RNAs that form
snap-back or hairpin structures (refs 2-7). This has the potential not only to allow stable
silencing of gene expression but also inducible silencing as has been observed in
trypanosomes and adult *Drosophila* (refs 2,4,5). The utility of this approach is somewhat
hampered by the difficulties that arise in the construction of bacterial plasmids containing
the long inverted repeats that are necessary to provoke silencing. In a recent report, it was
stated that more than 1,000 putative clones were screed to identify the desired construct
(ref 7).

The presence of hairpin structures often induces plasmid rearrangement, in part
due to the E. coli sbc proteins that recognize and cleave cruciform DNA structures (ref 8).
We have developed a method for the construction of hairpins that does not require cloning
of inverted repeats, per se. Instead, the fragment of the gene that is to be silenced is
cloned as a direct repeat, and the inversion is accomplished by treatment with a site-
specific recombinase, either *in vitro* (or potentially *in vivo)* (see Fig 29). Following
recombination, the inverted repeat structure is stable in a bacterial strain that lacks an
intact SBC system (DL759). We have successfully used this strategy to construct
numerous hairpin expression constructs that have been successfully used to provoke gene
silencing in Drosophila cells.

Literature Cited in Example 3

1.   Bosher, J. M. & Labouesse, M. RNA interference: genetic wand and genetic
      watchdog. *Nat Cell Biol* 2, E31-6 (2000).

2.   Fortier, E. & Belote, J. M. Temperature-dependent gene silencing by an expressed
      inverted repeat in Drosophila [published erratum appears in Genesis;2000
      May;27(1):47]. *Genesis* 26, 240-4 (2000).

3. Kennerdell, J. R. & Carthew, R. W. Heritable gene silencing in Drosophila using double-stranded RNA. *Nat Biotechnol* 18, 896-8 (2000).

4. Lam, G. & Thummel, C. S. Inducible expression of double-stranded RNA directs specific genetic interference in Drosophila [In Process Citation]. *Curr Biol* 10, 957-63 (2000).

5. Shi, H. *et al.* Genetic interference in Trypanosoma brucei by heritable and inducible double-stranded RNA. *Rna* 6, 1069-76 (2000).

6. Smith, N. A. *et al.* Total silencing by intron-spliced hairpin RNAs. *Nature* 407, 319-20 (2000).

7. Tavernarakis, N., Wang, S. L., Dorovkov, M., Ryazanov, A. & Driscoll, M. Heritable and inducible genetic interference by double-stranded RNA encoded by transgenes. *Nat Genet* 24, 180-3 (2000).

8. Connelly, J. C. & Leach, D. R. The sbcC and sbcD genes of Escherichia coli encode a nuclease involved in palindrome inviability and genetic recombination. *Genes Cells* 1, 285-91 (1996).

V. *Equivalents*

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

All of the above-cited references and publications are hereby incorporated by reference.

We Claim:

1. A method for attenuating expression of a target gene in a non-embryonic cell suspended in culture, comprising introducing into the cell a double stranded RNA (dsRNA) in an amount sufficient to attenuate expression of the target gene, wherein the dsRNA comprises a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence of the target gene.

2. A method for attenuating expression of a target gene in a mammalian cell, comprising

    (i)   activating one or both of a Dicer activity or an Argonaut activity in the cell, and

    (ii) introducing into the cell a double stranded RNA (dsRNA) in an amount sufficient to attenuate expression of the target gene, wherein the dsRNA comprises a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence of the target gene.

3. The method of claim 2, wherein the cell is suspended in culture.

4. The method of claim 2, wherein the cell is in a whole animal, such as a non-human mammal.

5. The method of claim 1 or 2, wherein is engineered with (i) a recombinant gene encoding a Dicer activity, (ii) a recombinant gene encoding an Argonaut activity, or (iii) both.

6. The method of claim 5, wherein the recombinant gene encodes a protein which includes an amino acid sequence at least 50 percent identical to SEQ ID No. 2 or 4 or the Argonaut sequence shown in Figure 24.

7. The method of claim 5, wherein the recombinant gene includes a coding sequence hybridizes under wash conditions of 2 x SSC at 22°C to SEQ ID No. 1 or 3.

8. The method of claim 1 or 2, wherein an endogenous Dicer gene or Argonaut gene is activated.

9. The method of claim 1 or 2, wherein the target gene is an endogenous gene of the cell.

10. The method of claim 1 or 2, wherein the target gene is an heterologous gene relative to the genome of the cell, such as a pathogen gene.

11. The method of claim 1 or 2, wherein the cell is treated with an agent that inhibits protein kinase RNA-activated (PKR) apoptosis, such as by treatment with agents which inhibit expression of PKR, cause its destruction, and/or inhibit the kinase activity of PKF.

12. The method of claim 1 or 2, wherein the cell is a primate cell, such as a human cell.

13. The method of claim 1 or 2, wherein the dsRNA is at least 50 nucleotides in length.

14. The method of claim 13, wherein the dsRNA is 400-800 nucleotides in length.

15. The method of claim 13, wherein the dsRNA is 400-800 nucleotides in length.

16. An assay for identifying nucleic acid sequences responsible for conferring a particular phenotype in a cell, comprising
    (i)   constructing a variegated library of nucleic acid sequences from a cell in an orientation relative to a promoter to produce double stranded DNA;
    (ii)   introducing the variegated dsRNA library into a culture of target cells, which cells have an activated Dicer activity or Argonaut activity;
    (iii) identifying members of the library which confer a particular phenotype on the cell, and identifying the sequence from a cell which correspond, such as being identical or homologous, to the library member.

17. A method of conducting a drug discovery business comprising:
    (i)   identifying, by the assay of claim 16, a target gene which provides a phenotypically desirable response when inhibited by RNAi;
    (ii)   identifying agents by their ability to inhibit expression of the target gene or the activity of an expression product of the target gene;
    (iii) conducting therapeutic profiling of agents identified in step (b), or further analogs thereof, for efficacy and toxicity in animals; and

(iv) formulating a pharmaceutical preparation including one or more agents identified in step (iii) as having an acceptable therapeutic profile.

18. The method of claim 17, including an additional step of establishing a distribution system for distributing the pharmaceutical preparation for sale, and may optionally

5 include establishing a sales group for marketing the pharmaceutical preparation.

19. A method of conducting a target discovery business comprising:

(i) identifying, by the assay of claim 16, a target gene which provides a phenotypically desirable response when inhibited by RNAi;

10 (ii) (optionally) conducting therapeutic profiling of the target gene for efficacy and toxicity in animals; and

(iii). licensing, to a third party, the rights for further drug development of inhibitors of the target gene.

15

a

No dsRNA   CD8 dsRNA   lacZ dsRNA

b dsRNA

None   lacZ   cyclin E

Ungated

Gated

c

dsRNA

Probe   *cyclin E*   *fizzy*   *cyclin A*

Figure 2

a

Figure 3

Figure 4

a

S100
Crude extract

~25 nt.—

b

lacZ

cyclin E

~25 nt.—

cyclin E
Northern

Figure 5

Figure 6a-c

Figure 6d-f

F.

marker
control
RISC
Dicer IP

E.

RISC - ls

RISC - hs

total

D.

ATP   IP   Ext

Figure 7

Figure 8A, B

Figure 9

Cell free extract

↓ *200K spin*

Ribosome pellet

→ *High salt extraction*

Ribosome associated proteins

Low salt precipitation

Supernatant

Pellet

RNAi activity

Resolubilize High Salt → Chromatography

Superose 6

Mono S

Mono Q

Hydroxyapatite

Figure 10

Figure 11

Fraction

Luciferase transcript

dArgo-2 immunoblot

216

129

91

Figure 12

Fraction

Luciferase transcript

dArgo-2 immunoblot

216
129
91

Figure 13

Fraction

Luciferase transcript

dArgo-2 immunoblot

Figure 14

Figure 15

Figure 16
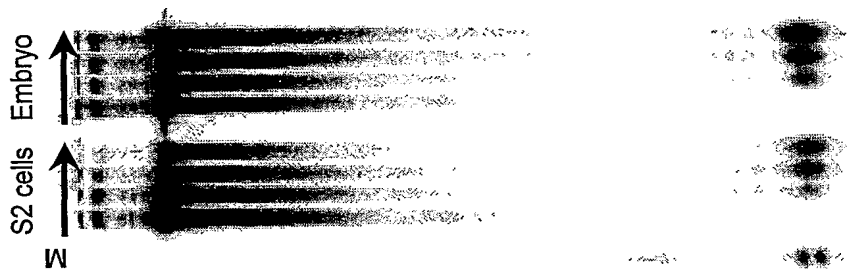
Figure 17
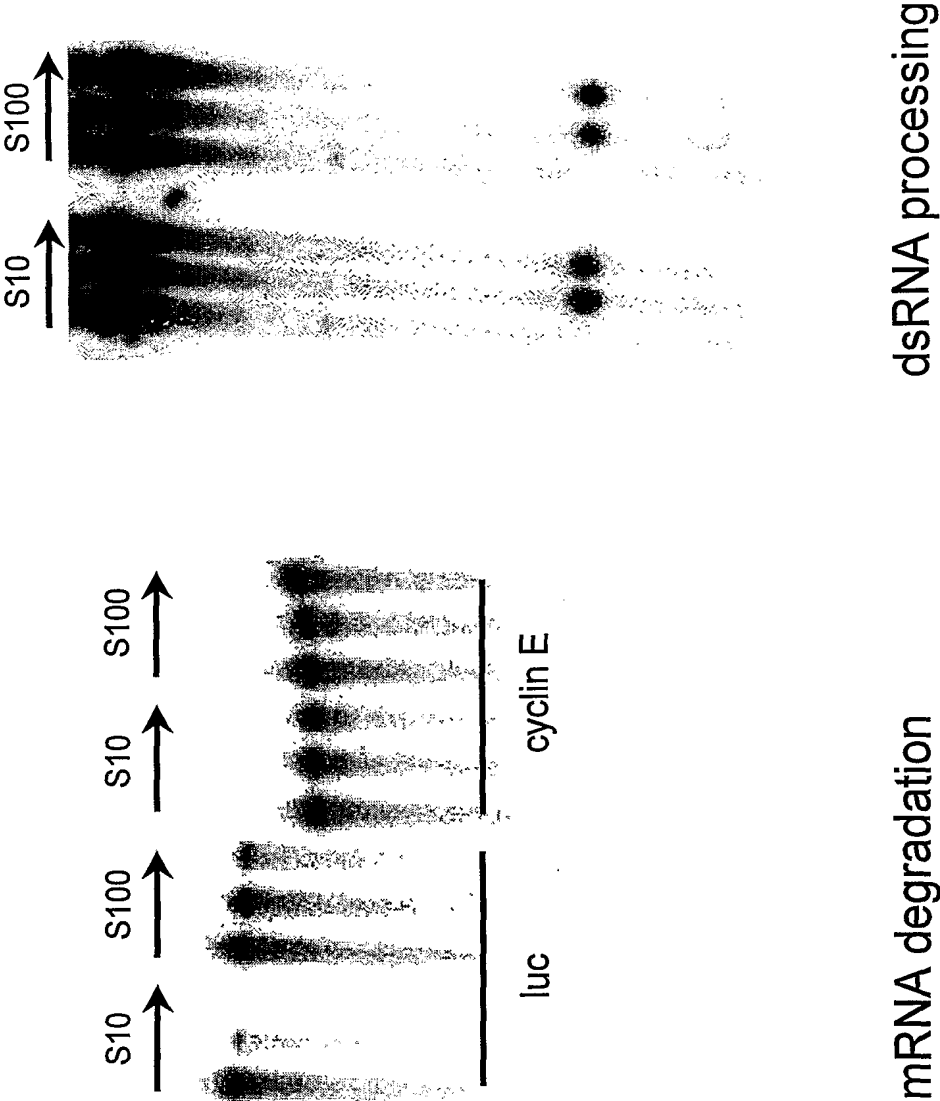
Figure 18
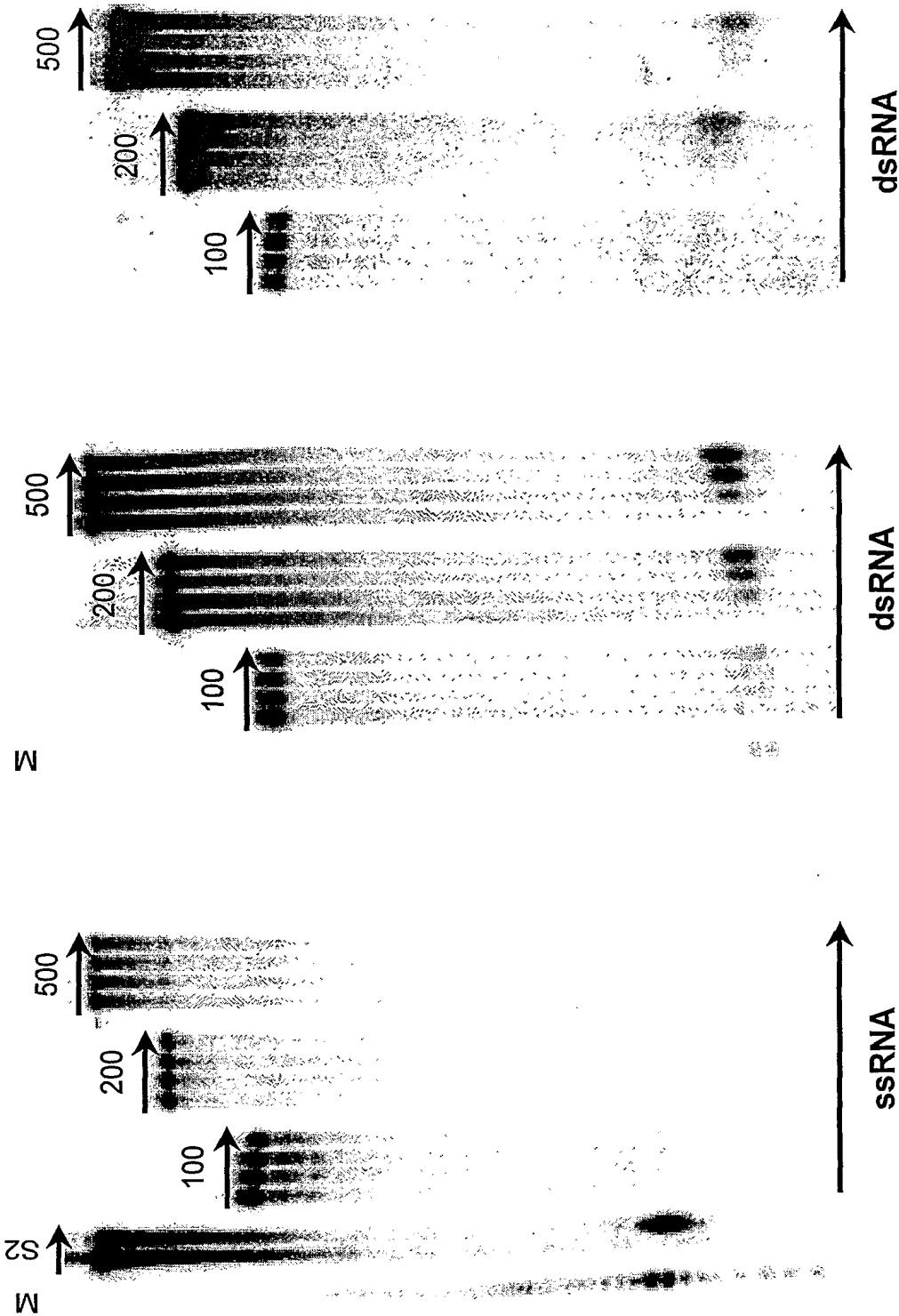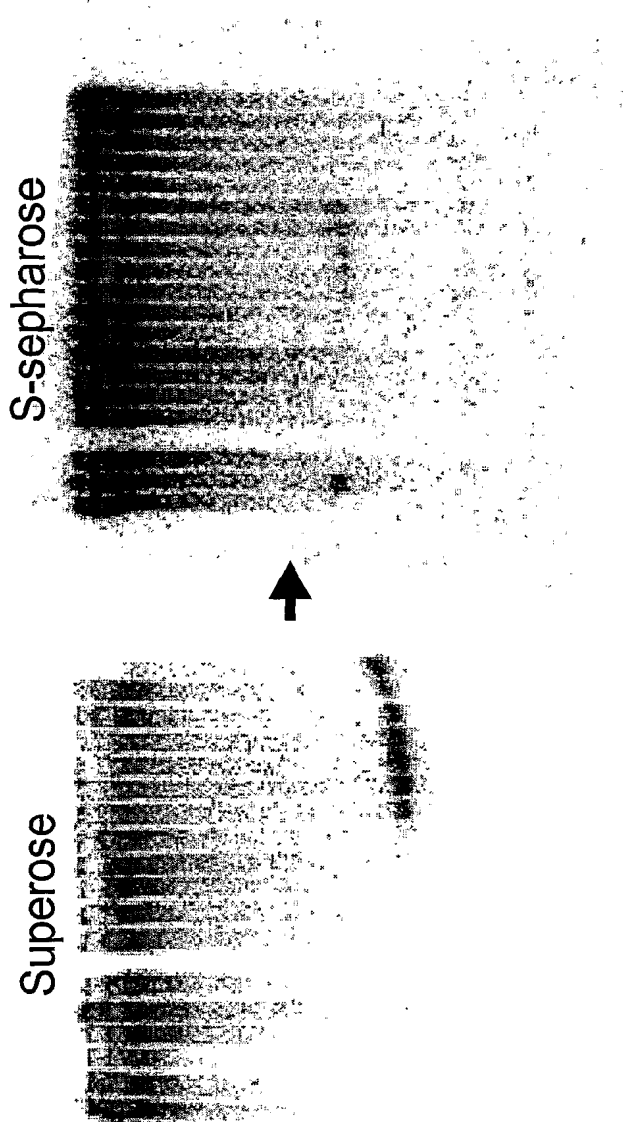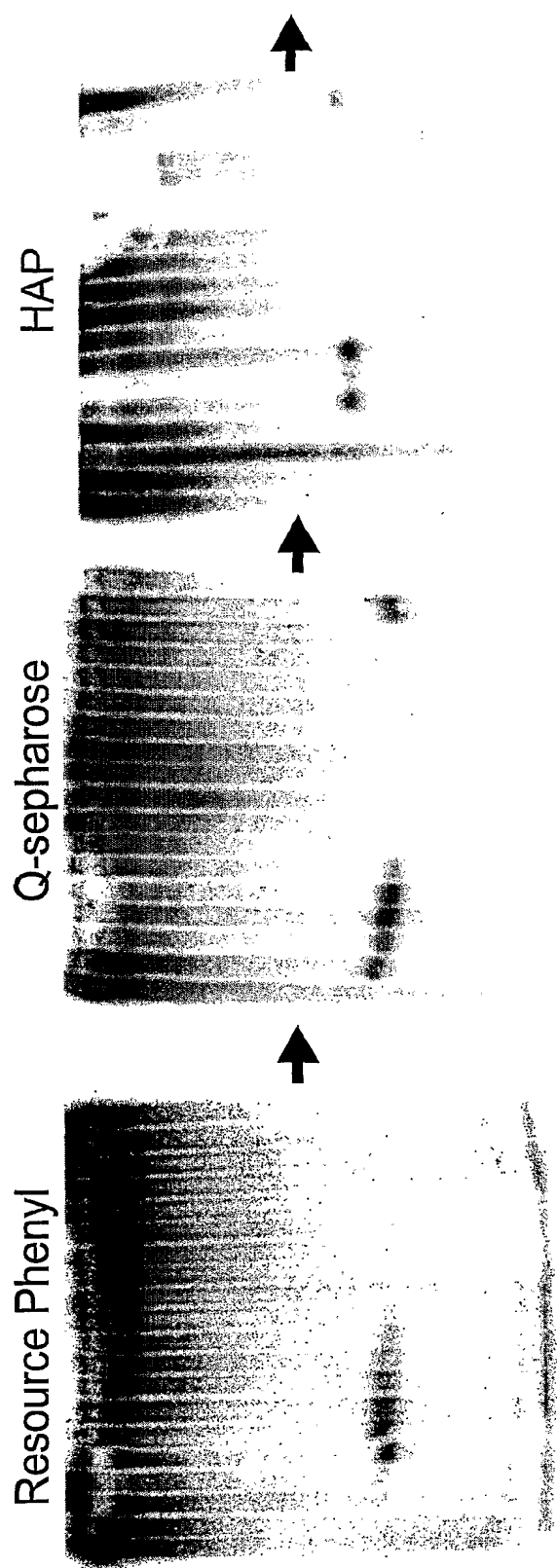
Figure 19

Purification of the 22-mer generating enzyme

Figure 20
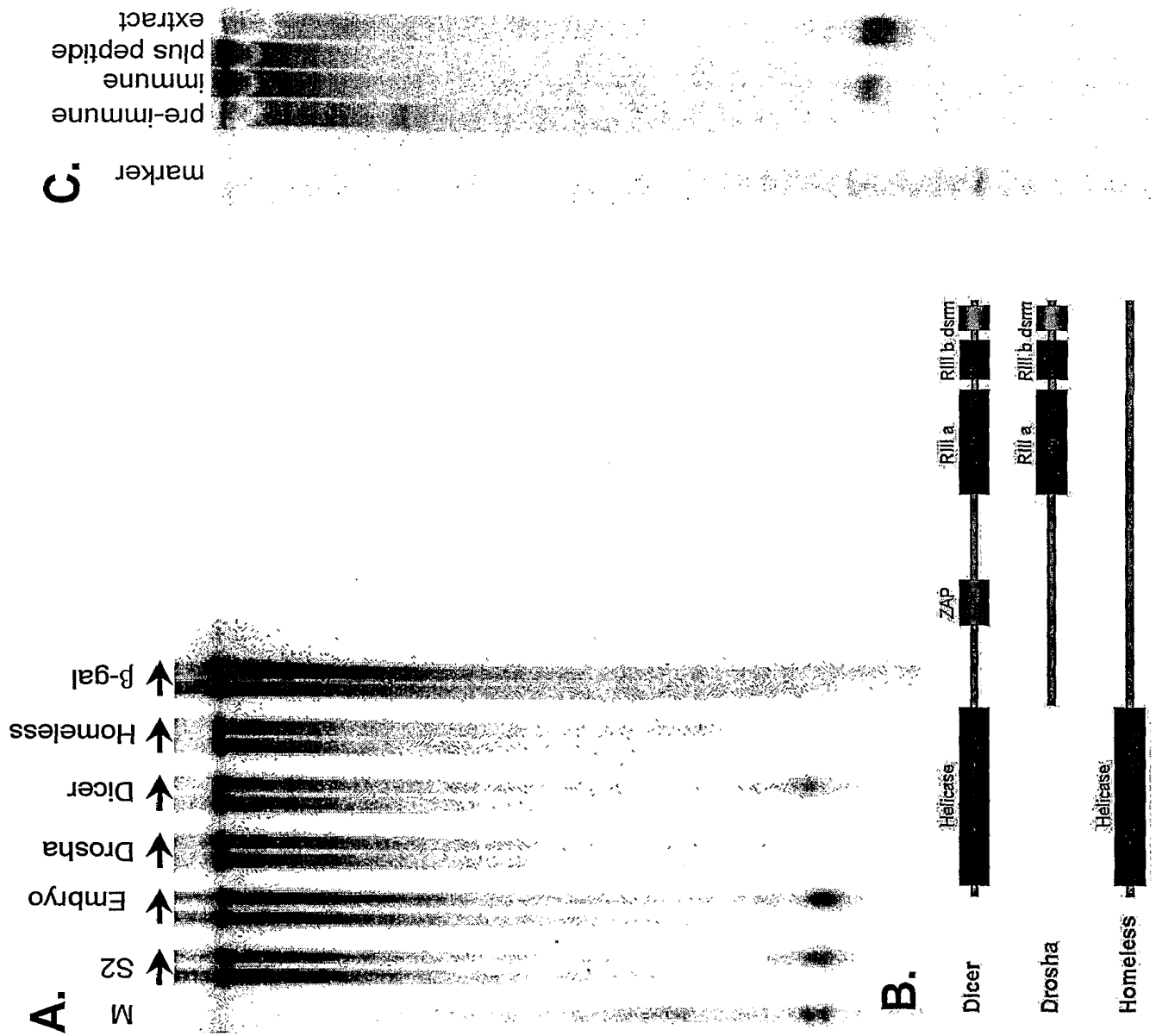
Figure 21
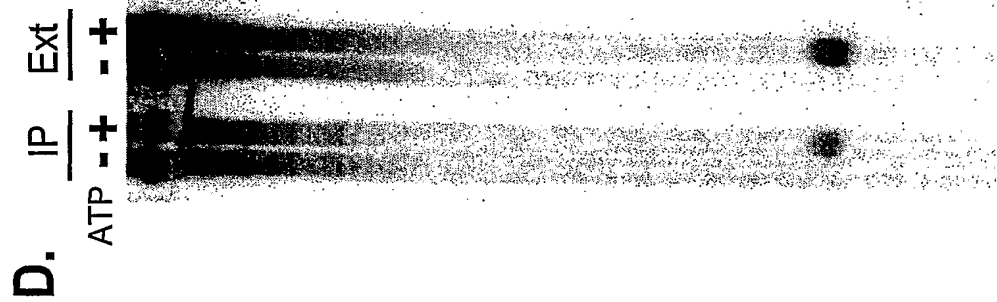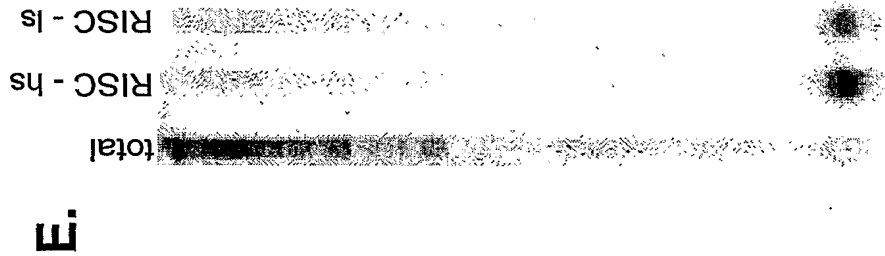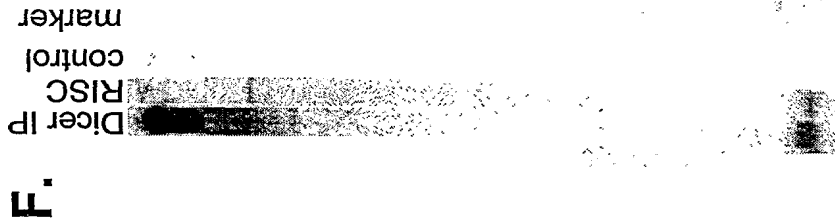
Figure 22
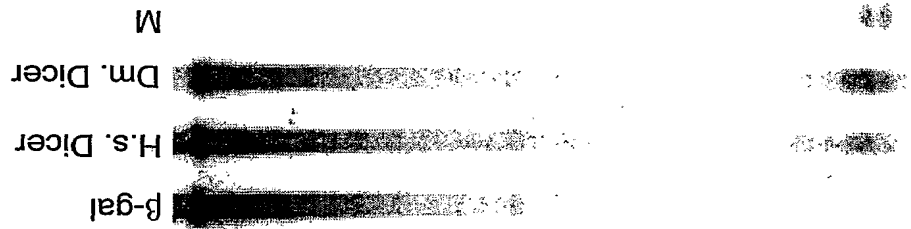
Figure 23

Figure 24

MGKKDKNKKGGGQDSAAAPQPQQQQKQQQQRQQQQPQQLQQPQQLQQPQQLQQPQQQQQQ
QPHQQQQQSRQQPSTSSGGSRASGFQQGGQQQKSQDAEGWTAQKKQGKQQVQGWTKQ
GQQGGHQQGRQGGRQGQGDGGYQQRPPGQQQGGHQQGRQGEGGYQQRPPGQQQGGHQQGRQG
QEGGYQQRPSGQQQGGHQQGRQGEGGYQQRPSGQQQGGHQQGRQGEGGYQQRPSGQ
QQGGHQQGRQGQEGGYQQRPSGQQQGGHQQGRQGEGGYQQRPSGQQQGGHQQGRQGQ
EGGYQQRPPGQQPNQTQSQGQYQSRGPPQQQAAPLPLPPQPAGSIKRGTIGKPGQVG
INYLDLDLSKMPSVAYHYDVKIMPERPKKFYRQAFEQFRVDQLGGAVLAYDGKASCYS
VDKLPLNSQNPEVTVTDRNGRTLRYTIEIKETGDSTIDLKSLTTYMNDRIFDKPMRAM
QCVEVVLASPCHNKAIRVGRSFFKMSDPNNRHELDDGYEALVGLYQAFMLGDRPFLNV
DISHKSFPISMPMIEYLERFSLKAKINNTTNLDYSRRFLEPFLRGINVVTPPQSFQS
APRVYRVNGLSRAPASSETFEHDGKKVTIASYFHSRNYPLKFPQLHCLNVGSSIKSIL
LPIELCSIEEGQALNRKDGATQVANMIKYAATSTNVRKRKIMNLLQYFQHNLDPTISR
FGIRIANDFIVVSTRVLSPPQVEYHSKRFTMVKNGSWRMDGMKFLEPKPKAHKCAVLY
CDPRSGRKMNYTQLNDFGNLIISQGKAVNISLDSDVTYRPFTDDERSLDTIFADLKRS
QHDLAIVIIPQFRISYDTIKQKAELQHGILTQCIKQFTVERKCNNQTIGNILLKINSK
LNGINHKIKDDPRLPMMKNTMYIGADVTHPSPDQREIPSVVGVAASHDPYGASYNMQY
RLQRGALEEIEDMFSITLEHLRVYKEYRNAYPDHIIYYRDGVSDGQFPKIKNEELRCI
KQACDKVGCKPKICCVIVVKRHHTRFFPSGDVTTSNKFNNVDPGTVVDRTIVHPNEMQ
FFMVSHQAIQGTAKPTRYNVIENTGNLDIDLLQQLTYNLCHMFPRCNRSVSYPAPAYL
AHLVAARGRVYLTGTNRFLDLKKEYAKRTIVPEEMKKNPMYFV
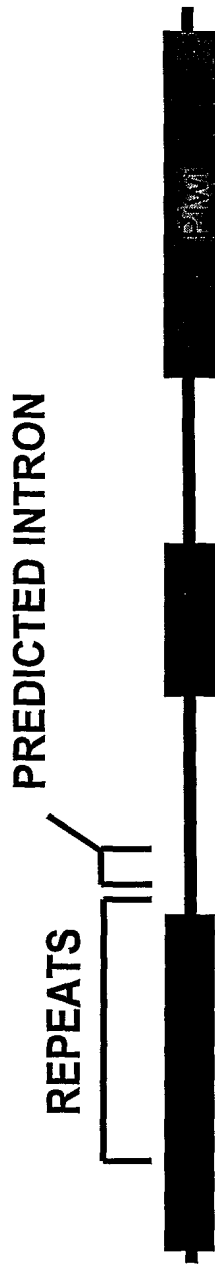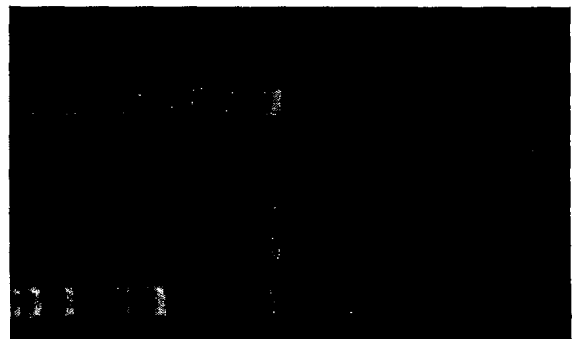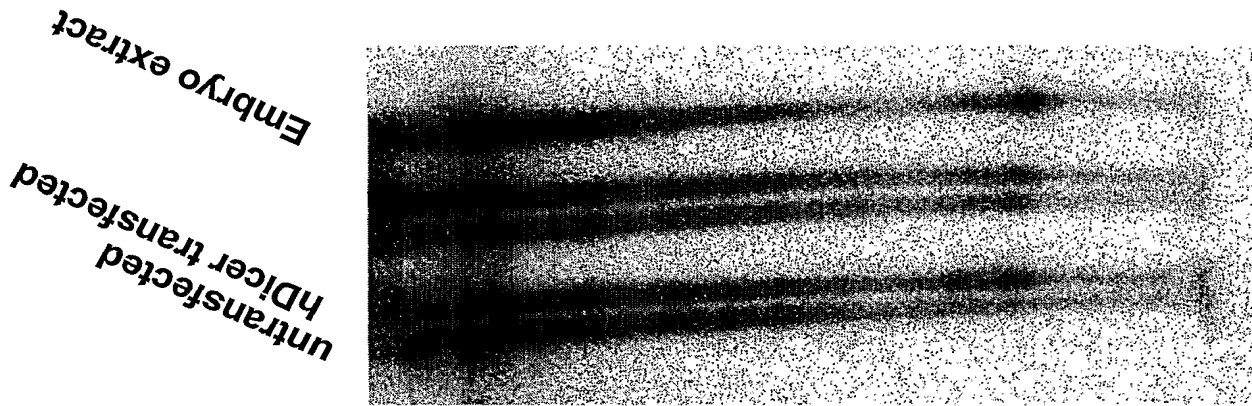
Figure 25

Figure 26

Figure 27
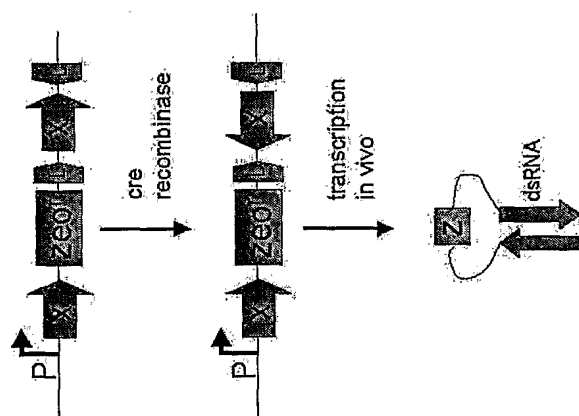
SEQUENCE LISTING

<110> Genetica, Inc.

<120> Methods and Compositions for RNA Interference

<130> GCNA-pWO-007

<140> PCT/US01/---,---
<141> 2001-03-16

<150> US 60/189,739
<151> 2000-03-16

<150> US 60/243,097
<151> 2000-10-24

<160> 4

<170> PatentIn version 3.0

<210> 1
<211> 5775
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(5775)

<400> 1
atg aaa agc cct gct ttg caa ccc ctc agc atg gca ggc ctg cag ctc   48
Met Lys Ser Pro Ala Leu Gln Pro Leu Ser Met Ala Gly Leu Gln Leu
1   5   10   15

atg acc cct gct tcc tca cca atg ggt cct ttc ttt gga ctg cca tgg   96
Met Thr Pro Ala Ser Ser Pro Met Gly Pro Phe Phe Gly Leu Pro Trp

               20   25   30

caa caa gaa gca att cat gat aac att tat acg cca aga aaa tat cag   144
Gln Gln Glu Ala Ile His Asp Asn Ile Tyr Thr Pro Arg Lys Tyr Gln
        35   40   45

gtt gaa ctg ctt gaa gca gct ctg gat cat aat acc atc gtc tgt tta   192
Val Glu Leu Leu Glu Ala Ala Leu Asp His Asn Thr Ile Val Cys Leu
        50   55   60

aac act ggc tca ggg aag aca ttt att gct agt act act cta cta aag   240
Asn Thr Gly Ser Gly Lys Thr Phe Ile Ala Ser Thr Thr Leu Leu Lys
65   70   75   80

agc tgt ctc tat cta gat cta ggg gag act tca gct aga aat gga aaa   288
Ser Cys Leu Tyr Leu Asp Leu Gly Glu Thr Ser Ala Arg Asn Gly Lys
                85   90   95

agg acg gtg ttc ttg gtc aac tct gca aac cag gtt gct caa caa gtg   336
Arg Thr Val Phe Leu Val Asn Ser Ala Asn Gln Val Ala Gln Gln Val
        100   105   110

tca gct gtc aga act cat tca gat ctc aag gtt ggg gaa tac tca aac   384
Ser Ala Val Arg Thr His Ser Asp Leu Lys Val Gly Glu Tyr Ser Asn
        115   120   125

cta gaa gta aat gca tct tgg aca aaa gag aga tgg aac caa gag ttt   432
Leu Glu Val Asn Ala Ser Trp Thr Lys Glu Arg Trp Asn Gln Glu Phe
        130   135   140

act aag cac cag gtt ctc att atg act tgc tat gtc gcc ttg aat gtt   480
Thr Lys His Gln Val Leu Ile Met Thr Cys Tyr Val Ala Leu Asn Val
145   150   155   160

ttg aaa aat ggt tac tta tca ctg tca gac att aac ctt ttg gtg ttt   528
Leu Lys Asn Gly Tyr Leu Ser Leu Ser Asp Ile Asn Leu Leu Val Phe
        165   170   175

gat gag tgt cat ctt gca atc cta gac cac ccc tat cga gaa ttt atg   576
Asp Glu Cys His Leu Ala Ile Leu Asp His Pro Tyr Arg Glu Phe Met
           180   185   190

aag ctc tgt gaa att tgt cca tca tgt cct cgc att ttg gga cta act   624
Lys Leu Cys Glu Ile Cys Pro Ser Cys Pro Arg Ile Leu Gly Leu Thr
            195   200   205

gct tcc att tta aat ggg aaa tgg gat cca gag gat ttg gaa gaa aag   672
Ala Ser Ile Leu Asn Gly Lys Trp Asp Pro Glu Asp Leu Glu Glu Lys
           210   215   220

ttt cag aaa cta gag aaa att ctt aag agt aat gct gaa act gca act   720
Phe Gln Lys Leu Glu Lys Ile Leu Lys Ser Asn Ala Glu Thr Ala Thr
225   230   235   240

gac ctg gtg gtc tta gac agg tat act tct cag cca tgt gag att gtg   768
Asp Leu Val Val Leu Asp Arg Tyr Thr Ser Gln Pro Cys Glu Ile Val
                245   250   255

gtg gat tgt gga cca ttt act gac aga agt ggg ctt tat gaa aga ctg   816
Val Asp Cys Gly Pro Phe Thr Asp Arg Ser Gly Leu Tyr Glu Arg Leu
           260   265   270

ctg atg gaa tta gaa gaa gca ctt aat ttt atc aat gat tgt aat ata   864
Leu Met Glu Leu Glu Glu Ala Leu Asn Phe Ile Asn Asp Cys Asn Ile
           275   280   285

tct gta cat tca aaa gaa aga gat tct act tta att tcg aaa cag ata   912
Ser Val His Ser Lys Glu Arg Asp Ser Thr Leu Ile Ser Lys Gln Ile
           290   295   300

cta tca gac tgt cgt gcc gta ttg gta gtt ctg gga ccc tgg tgt gca   960
Leu Ser Asp Cys Arg Ala Val Leu Val Val Leu Gly Pro Trp Cys Ala
305   310   315   320

3

gat aaa gta gct gga atg atg gta aga gaa cta cag aaa tac atc aaa   1008
Asp Lys Val Ala Gly Met Met Val Arg Glu Leu Gln Lys Tyr Ile Lys
            325   330   335

cat gag caa gag gag ctg cac agg aaa ttt tta ttg ttt aca gac act   1056
His Glu Gln Glu Glu Leu His Arg Lys Phe Leu Leu Phe Thr Asp Thr
            340   345   350

ttc cta agg aaa ata cat gca cta tgt gaa gag cac ttc tca cct gcc   1104
Phe Leu Arg Lys Ile His Ala Leu Cys Glu Glu His Phe Ser Pro Ala
            355   360   365

tca ctt gac ctg aaa ttt gta act cct aaa gta atc aaa ctg ctc gaa   1152
Ser Leu Asp Leu Lys Phe Val Thr Pro Lys Val Ile Lys Leu Leu Glu
          370   375   380

atc tta cgc aaa tat aaa cca tat gag cga cac agt ttt gaa agc gtt   1200
Ile Leu Arg Lys Tyr Lys Pro Tyr Glu Arg His Ser Phe Glu Ser Val
385   390   395   400

gag tgg tat aat aat aga aat cag gat aat tat gtg tca tgg agt gat   1248
Glu Trp Tyr Asn Asn Arg Asn Gln Asp Asn Tyr Val Ser Trp Ser Asp
              405   410   415

tct gag gat gat gat gag gat gaa gaa att gaa gaa aaa gag aag cca   1296
Ser Glu Asp Asp Asp Glu Asp Glu Glu Ile Glu Glu Lys Glu Lys Pro
              420   425   430

gag aca aat ttt cct tct cct ttt acc aac att ttg tgc gga att att   1344
Glu Thr Asn Phe Pro Ser Pro Phe Thr Asn Ile Leu Cys Gly Ile Ile
            435   440   445

ttt gtg gaa aga aga tac aca gca gtt gtc tta aac aga ttg ata aag   1392
Phe Val Glu Arg Arg Tyr Thr Ala Val Val Leu Asn Arg Leu Ile Lys
            450   455   460

gaa gct ggc aaa caa gat cca gag ctg gct tat atc agt agc aat ttc   1440

Glu Ala Gly Lys Gln Asp Pro Glu Leu Ala Tyr Ile Ser Ser Asn Phe
465   470   475   480

ata act gga cat ggc att ggg aag aat cag cct cgc aac aac acg atg   1488
Ile Thr Gly His Gly Ile Gly Lys Asn Gln Pro Arg Asn Asn Thr Met
          485   490   495

gaa gca gaa ttc aga aaa cag gaa gag gta ctt agg aaa ttt cga gca   1536
Glu Ala Glu Phe Arg Lys Gln Glu Glu Val Leu Arg Lys Phe Arg Ala
          500   505   510

cat gag acc aac ctg ctt att gca aca agt att gta gaa gag ggt gtt   1584
His Glu Thr Asn Leu Leu Ile Ala Thr Ser Ile Val Glu Glu Gly Val
          515   520   525

gat ata cca aaa tgc aac ttg gtg gtt cgt ttt gat ttg ccc aca gaa   1632
Asp Ile Pro Lys Cys Asn Leu Val Val Arg Phe Asp Leu Pro Thr Glu
     530   535   540

tat cga tcc tat gtt caa tct aaa gga aga gca agg gca ccc atc tct   1680
Tyr Arg Ser Tyr Val Gln Ser Lys Gly Arg Ala Arg Ala Pro Ile Ser
545   550   555   560

aat tat ata atg tta gcg gat aca gac aaa ata aaa agt ttt gaa gaa   1728
Asn Tyr Ile Met Leu Ala Asp Thr Asp Lys Ile Lys Ser Phe Glu Glu
          565   570   575

gac ctt aaa acc tac aaa gct att gaa aag atc ttg aga aac aag tgt   1776
Asp Leu Lys Thr Tyr Lys Ala Ile Glu Lys Ile Leu Arg Asn Lys Cys
          580   585   590

tcc aag tcg gtt gat act ggt gag act gac att gat cct gtc atg gat   1824
Ser Lys Ser Val Asp Thr Gly Glu Thr Asp Ile Asp Pro Val Met Asp
          595   600   605

gat gat cac gtt ttc cca cca tat gtg ttg agg cct gac gat ggt ggt   1872
Asp Asp His Val Phe Pro Pro Tyr Val Leu Arg Pro Asp Asp Gly Gly

```
        610            615            620

cca cga gtc aca atc aac acg gcc att gga cac atc aat aga tac tgt    1920
Pro Arg Val Thr Ile Asn Thr Ala Ile Gly His Ile Asn Arg Tyr Cys
625            630            635            640


gct aga tta cca agt gat ccg ttt act cat cta gct cct aaa tgc aga    1968
Ala Arg Leu Pro Ser Asp Pro Phe Thr His Leu Ala Pro Lys Cys Arg
           645            650            655


acc cga gag ttg cct gat ggt aca ttt tat tca act ctt tat ctg cca    2016
Thr Arg Glu Leu Pro Asp Gly Thr Phe Tyr Ser Thr Leu Tyr Leu Pro
           660            665            670


att aac tca cct ctt cga gcc tcc att gtt ggt cca cca atg agc tgt    2064
Ile Asn Ser Pro Leu Arg Ala Ser Ile Val Gly Pro Pro Met Ser Cys
           675            680            685


gta cga ttg gct gaa aga gtt gtc gct ctc att tgc tgt gag aaa ctg    2112
Val Arg Leu Ala Glu Arg Val Val Ala Leu Ile Cys Cys Glu Lys Leu
           690            695            700


cac aaa att ggc gaa ctg gat gac cat ttg atg cca gtt ggg aaa gag    2160
His Lys Ile Gly Glu Leu Asp Asp His Leu Met Pro Val Gly Lys Glu
705            710            715            720


act gtt aaa tat gaa gag gag ctt gat ttg cat gat gaa gaa gag acc    2208
Thr Val Lys Tyr Glu Glu Glu Leu Asp Leu His Asp Glu Glu Glu Thr
           725            730            735


agt gtt cca gga aga cca ggt tcc acg aaa cga agg cag tgc tac cca    2256
Ser Val Pro Gly Arg Pro Gly Ser Thr Lys Arg Arg Gln Cys Tyr Pro
           740            745            750


aaa gca att cca gag tgt ttg agg gat agt tat ccc aga cct gat cag    2304
Lys Ala Ile Pro Glu Cys Leu Arg Asp Ser Tyr Pro Arg Pro Asp Gln
           755            760            765
```

```
ccc tgt tac ctg tat gtg ata gga atg gtt tta act aca cct tta cct     2352
Pro Cys Tyr Leu Tyr Val Ile Gly Met Val Leu Thr Thr Pro Leu Pro
     770            775            780


gat gaa ctc aac ttt aga agg cgg aag ctc tat cct cct gaa gat acc     2400
Asp Glu Leu Asn Phe Arg Arg Arg Lys Leu Tyr Pro Pro Glu Asp Thr
     785            790            795            800


aca aga tgc ttt gga ata ctg acg gcc aaa ccc ata cct cag att cca     2448
Thr Arg Cys Phe Gly Ile Leu Thr Ala Lys Pro Ile Pro Gln Ile Pro
            805            810            815


cac ttt cct gtg tac aca cgc tct gga gag gtt acc ata tcc att gag     2496
His Phe Pro Val Tyr Thr Arg Ser Gly Glu Val Thr Ile Ser Ile Glu
            820            825            830


ttg aag aag tct ggt ttc atg ttg tct cta caa atg ctt gag ttg att     2544
Leu Lys Lys Ser Gly Phe Met Leu Ser Leu Gln Met Leu Glu Leu Ile
            835            840            845


aca aga ctt cac cag tat ata ttc tca cat att ctt cgg ctt gaa aaa     2592
Thr Arg Leu His Gln Tyr Ile Phe Ser His Ile Leu Arg Leu Glu Lys
     850            855            860


cct gca cta gaa ttt aaa cct aca gac gct gat tca gca tac tgt gtt     2640
Pro Ala Leu Glu Phe Lys Pro Thr Asp Ala Asp Ser Ala Tyr Cys Val
865            870            875            880


cta cct ctt aat gtt gtt aat gac tcc agc act ttg gat att gac ttt     2688
Leu Pro Leu Asn Val Val Asn Asp Ser Ser Thr Leu Asp Ile Asp Phe
            885            890            895


aaa ttc atg gaa gat att gag aag tct gaa gct cgc ata ggc att ccc     2736
Lys Phe Met Glu Asp Ile Glu Lys Ser Glu Ala Arg Ile Gly Ile Pro
            900            905            910
```

7

agt aca aag tat aca aaa gaa aca ccc ttt gtt ttt aaa tta gaa gat   2784
Ser Thr Lys Tyr Thr Lys Glu Thr Pro Phe Val Phe Lys Leu Glu Asp
        915   920   925

tac caa gat gcc gtt atc att cca aga tat cgc aat ttt gat cag cct   2832
Tyr Gln Asp Ala Val Ile Ile Pro Arg Tyr Arg Asn Phe Asp Gln Pro
        930   935   940

cat cga ttt tat gta gct gat gtg tac act gat ctt acc cca ctc agt   2880
His Arg Phe Tyr Val Ala Asp Val Tyr Thr Asp Leu Thr Pro Leu Ser
945   950   955   960

aaa ttt cct tcc cct gag tat gaa act ttt gca gaa tat tat aaa aca   2928
Lys Phe Pro Ser Pro Glu Tyr Glu Thr Phe Ala Glu Tyr Tyr Lys Thr
            965   970   975

aag tac aac ctt gac cta acc aat ctc aac cag cca ctg ctg gat gtg   2976
Lys Tyr Asn Leu Asp Leu Thr Asn Leu Asn Gln Pro Leu Leu Asp Val
            980   985   990

gac cac aca tct tca aga ctt aat ctt ttg aca cct cga cat ttg aat   3024
Asp His Thr Ser Ser Arg Leu Asn Leu Leu Thr Pro Arg His Leu Asn
            995   1000   1005

cag aag ggg aaa gcg ctt cct tta agc agt gct gag aag agg aaa   3069
Gln Lys Gly Lys Ala Leu Pro Leu Ser Ser Ala Glu Lys Arg Lys
        1010   1015   1020

gcc aaa tgg gaa agt ctg cag aat aaa cag ata ctg gtt cca gaa   3114
Ala Lys Trp Glu Ser Leu Gln Asn Lys Gln Ile Leu Val Pro Glu
        1025   1030   1035

ctc tgt gct ata cat cca att cca gca tca ctg tgg aga aaa gct   3159
Leu Cys Ala Ile His Pro Ile Pro Ala Ser Leu Trp Arg Lys Ala
        1040   1045   1050

gtt tgt ctc ccc agc ata ctt tat cgc ctt cac tgc ctt ttg act   3204

Val Cys Leu Pro Ser Ile Leu Tyr Arg Leu His Cys Leu Leu Thr
  1055   1060   1065

gca gag gag cta aga gcc cag act gcc agc gat gct ggc gtg gga   3249
Ala Glu Glu Leu Arg Ala Gln Thr Ala Ser Asp Ala Gly Val Gly
  1070   1075   1080

gtc aga tca ctt cct gcg gat ttt aga tac cct aac tta gac ttc   3294
Val Arg Ser Leu Pro Ala Asp Phe Arg Tyr Pro Asn Leu Asp Phe
  1085   1090   1095

ggg tgg aaa aaa tct att gac agc aaa tct ttc atc tca att tct   3339
Gly Trp Lys Lys Ser Ile Asp Ser Lys Ser Phe Ile Ser Ile Ser
  1100   1105   1110

aac tcc tct tca gct gaa aat gat aat tac tgt aag cac agc aca   3384
Asn Ser Ser Ser Ala Glu Asn Asp Asn Tyr Cys Lys His Ser Thr
  1115   1120   1125

att gtc cct gaa aat gct gca cat caa ggt gct aat .aga acc tcc   3429
Ile Val Pro Glu Asn Ala Ala His Gln Gly Ala Asn Arg Thr Ser
  1130   1135   1140

tct cta gaa aat cat gac caa atg tct gtg aac tgc aga acg ttg   3474
Ser Leu Glu Asn His Asp Gln Met Ser Val Asn Cys Arg Thr Leu
  1145   1150   1155

ctc agc gag tcc cct ggt aag ctc cac gtt gaa gtt tca gca gat   3519
Leu Ser Glu Ser Pro Gly Lys Leu His Val Glu Val Ser Ala Asp
  1160   1165   1170

ctt aca gca att aat ggt ctt tct tac aat caa aat ctc gcc aat   3564
Leu Thr Ala Ile Asn Gly Leu Ser Tyr Asn Gln Asn Leu Ala Asn
  1175   1180   1185

ggc agt tat gat tta gct aac aga gac ttt tgc caa gga aat cag · 3609
Gly Ser Tyr Asp Leu Ala Asn Arg Asp Phe Cys Gln Gly Asn Gln

1190   1195   1200

cta aat tac tac aag cag gaa ata ccc gtg caa cca act acc tca   3654
Leu Asn Tyr Tyr Lys Gln Glu Ile Pro Val Gln Pro Thr Thr Ser
    1205   1210   1215

tat tcc att cag aat tta tac agt tac gag aac cag ccc cag ccc   3699
Tyr Ser Ile Gln Asn Leu Tyr Ser Tyr Glu Asn Gln Pro Gln Pro
    1220   1225   1230

agc gat gaa tgt act ctc ctg agt aat aaa tac ctt gat gga aat   3744
Ser Asp Glu Cys Thr Leu Leu Ser Asn Lys Tyr Leu Asp Gly Asn
    1235   1240   1245

gct aac aaa tct acc tca gat gga agt cct gtg atg gcc gta atg   3789
Ala Asn Lys Ser Thr Ser Asp Gly Ser Pro Val Met Ala Val Met
    1250   1255   1260

cct ggt acg aca gac act att caa gtg ctc aag ggc agg atg gat   3834
Pro Gly Thr Thr Asp Thr Ile Gln Val Leu Lys Gly Arg Met Asp
    1265   1270   1275

tct gag cag agc cct tct att ggg tac tcc tca agg act ctt ggc   3879
Ser Glu Gln Ser Pro Ser Ile Gly Tyr Ser Ser Arg Thr Leu Gly
    1280   1285   1290

ccc aat cct gga ctt att ctt cag gct ttg act ctg tca aac gct   3924
Pro Asn Pro Gly Leu Ile Leu Gln Ala Leu Thr Leu Ser Asn Ala
    1295   1300   1305

agt gat gga ttt aac ctg gag cgg ctt gaa atg ctt ggc gac tcc   3969
Ser Asp Gly Phe Asn Leu Glu Arg Leu Glu Met Leu Gly Asp Ser
    1310   1315   1320

ttt tta aag cat gcc atc acc aca tat cta ttt tgc act tac cct   4014
Phe Leu Lys His Ala Ile Thr Thr Tyr Leu Phe Cys Thr Tyr Pro
    1325   1330   1335

gat gcg cat gag ggc cgc ctt tca tat atg aga agc aaa aag gtc   4059
Asp Ala His Glu Gly Arg Leu Ser Tyr Met Arg Ser Lys Lys Val
     1340   1345   1350

agc aac tgt aat ctg tat cgc ctt gga aaa aag aag gga cta ccc   4104
Ser Asn Cys Asn Leu Tyr Arg Leu Gly Lys Lys Lys Gly Leu Pro
     1355   1360   1365

agc cgc atg gtg gtg tca ata ttt gat ccc cct gtg aat tgg ctt   4149
Ser Arg Met Val Val Ser Ile Phe Asp Pro Pro Val Asn Trp Leu
     1370   1375   1380

cct cct ggt tat gta gta aat caa gac aaa agc aac aca gat aaa   4194
Pro Pro Gly Tyr Val Val Asn Gln Asp Lys Ser Asn Thr Asp Lys
     1385   1390   1395

tgg gaa aaa gat gaa atg aca aaa gac tgc atg ctg gcg aat ggc   4239
Trp Glu Lys Asp Glu Met Thr Lys Asp Cys Met Leu Ala Asn Gly
     1400   1405   1410

aaa ctg gat gag gat tac gag gag gag gat gag gag gag gag agc   4284
Lys Leu Asp Glu Asp Tyr Glu Glu Glu Asp Glu Glu Glu Glu Ser
     1415   1420   1425

ctg atg tgg agg gct ccg aag gaa gag gct gac tat gaa gat gat   4329
Leu Met Trp Arg Ala Pro Lys Glu Glu Ala Asp Tyr Glu Asp Asp
     1430   1435   1440

ttc ctg gag tat gat cag gaa cat atc aga ttt ata gat aat atg   4374
Phe Leu Glu Tyr Asp Gln Glu His Ile Arg Phe Ile Asp Asn Met
     1445   1450   1455

tta atg ggg tca gga gct ttt gta aag aaa atc tct ctt tct cct   4419
Leu Met Gly Ser Gly Ala Phe Val Lys Lys Ile Ser Leu Ser Pro
     1460   1465   1470

ttt tca acc act gat tct gca tat gaa tgg aaa atg ccc aaa aaa   4464
Phe Ser Thr Thr Asp Ser Ala Tyr Glu Trp Lys Met Pro Lys Lys
    1475   1480   1485

tcc tcc tta ggt agt atg cca ttt tca tca gat ttt gag gat ttt   4509
Ser Ser Leu Gly Ser Met Pro Phe Ser Ser Asp Phe Glu Asp Phe
    1490   1495   1500

gac tac agc tct tgg gat gca atg tgc tat ctg gat cct agc aaa   4554
Asp Tyr Ser Ser Trp Asp Ala Met Cys Tyr Leu Asp Pro Ser Lys
    1505   1510   1515

gct gtt gaa gaa gat gac ttt gtg gtg ggg ttc tgg aat cca tca   4599
Ala Val Glu Glu Asp Asp Phe Val Val Gly Phe Trp Asn Pro Ser
    1520   1525   1530

gaa gaa aac tgt ggt gtt gac acg gga aag cag tcc att tct tac   4644
Glu Glu Asn Cys Gly Val Asp Thr Gly Lys Gln Ser Ile Ser Tyr
    1535   1540   1545

gac ttg cac act gag cag tgt att gct gac aaa agc ata gcg gac   4689
Asp Leu His Thr Glu Gln Cys Ile Ala Asp Lys Ser Ile Ala Asp
    1550   1555   1560

tgt gtg gaa gcc ctg ctg ggc tgc tat tta acc agc tgt ggg gag   4734
Cys Val Glu Ala Leu Leu Gly Cys Tyr Leu Thr Ser Cys Gly Glu
    1565   1570   1575

agg gct gct cag ctt ttc ctc tgt tca ctg ggg ctg aag gtg ctc   4779
Arg Ala Ala Gln Leu Phe Leu Cys Ser Leu Gly Leu Lys Val Leu
    1580   1585   1590

ccg gta att aaa agg act gat cgg gaa aag gcc ctg tgc cct act   4824
Pro Val Ile Lys Arg Thr Asp Arg Glu Lys Ala Leu Cys Pro Thr
    1595   1600   1605

cgg gag aat ttc aac agc caa caa aag aac ctt tca gtg agc tgt   4869

Arg Glu Asn Phe Asn Ser Gln Gln Lys Asn Leu Ser Val Ser Cys
   1610   1615   1620

gct gct gct tct gtg gcc agt tca cgc tct tct gta ttg aaa gac   4914
Ala Ala Ala Ser Val Ala Ser Ser Arg Ser Ser Val Leu Lys Asp
   1625   1630   1635

tcg gaa tat ggt tgt ttg aag att cca cca aga tgt atg ttt gat   4959
Ser Glu Tyr Gly Cys Leu Lys Ile Pro Pro Arg Cys Met Phe Asp
   1640   1645   1650

cat cca gat gca gat aaa aca ctg aat cac ctt ata tcg ggg ttt   5004
His Pro Asp Ala Asp Lys Thr Leu Asn His Leu Ile Ser Gly Phe
   1655   1660   1665

gaa aat ttt gaa aag aaa atc aac tac aga ttc aag aat aag gct   5049
Glu Asn Phe Glu Lys Lys Ile Asn Tyr Arg Phe Lys Asn Lys Ala
   1670   1675   1680

tac ctt ctc cag gct ttt aca cat gcc tcc tac cac tac aat act   5094
Tyr Leu Leu Gln Ala Phe Thr His Ala Ser Tyr His Tyr Asn Thr
   1685   1690   1695

atc act gat tgt tac cag cgc tta gaa ttc ctg gga gat gcg att   5139
Ile Thr Asp Cys Tyr Gln Arg Leu Glu Phe Leu Gly Asp Ala Ile
   1700   1705   1710

ttg gac tac ctc ata acc aag cac ctt tat gaa gac ccg cgg cag   5184
Leu Asp Tyr Leu Ile Thr Lys His Leu Tyr Glu Asp Pro Arg Gln
   1715   1720   1725

cac tcc ccg ggg gtc ctg aca gac ctg cgg tct gcc ctg gtc aac   5229
His Ser Pro Gly Val Leu Thr Asp Leu Arg Ser Ala Leu Val Asn
   1730   1735   1740

aac acc atc ttt gca tcg ctg gct gta aag tac gac tac cac aag   5274
Asn Thr Ile Phe Ala Ser Leu Ala Val Lys Tyr Asp Tyr His Lys

        1745   1750   1755

tac ttc aaa gct gtc tct cct gag ctc ttc cat gtc att gat gac   5319
Tyr Phe Lys Ala Val Ser Pro Glu Leu Phe His Val Ile Asp Asp
     1760   1765   1770

ttt gtg cag ttt cag ctt gag aag aat gaa atg caa gga atg gat   5364
Phe Val Gln Phe Gln Leu Glu Lys Asn Glu Met Gln Gly Met Asp
     1775   1780   1785

tct gag ctt agg aga tct gag gag gat gaa gag aaa gaa gag gat   5409
Ser Glu Leu Arg Arg Ser Glu Glu Asp Glu Glu Lys Glu Glu Asp
     1790   1795   1800

att gaa gtt cca aag gcc atg ggg gat att ttt gag tcg ctt gct   5454
Ile Glu Val Pro Lys Ala Met Gly Asp Ile Phe Glu Ser Leu Ala
     1805   1810   1815

ggt gcc att tac atg gat agt ggg atg tca ctg gag aca gtc tgg   5499
Gly Ala Ile Tyr Met Asp Ser Gly Met Ser Leu Glu Thr Val Trp
     1820   1825   1830

cag gtg tac tat ccc atg atg cgg cca cta ata gaa aag ttt tct   5544
Gln Val Tyr Tyr Pro Met Met Arg Pro Leu Ile Glu Lys Phe Ser
     1835   1840   1845

gca aat gta ccc cgt tcc cct gtg cga gaa ttg ctt gaa atg gaa   5589
Ala Asn Val Pro Arg Ser Pro Val Arg Glu Leu Leu Glu Met Glu
     1850   1855   1860

cca gaa act gcc aaa ttt agc ccg gct gag aga act tac gac ggg   5634
Pro Glu Thr Ala Lys Phe Ser Pro Ala Glu Arg Thr Tyr Asp Gly
     1865   1870   1875

aag gtc aga gtc act gtg gaa gta gta gga aag ggg aaa ttt aaa   . 5679
Lys Val Arg Val Thr Val Glu Val Val Gly Lys Gly Lys Phe Lys
     1880   1885   1890

ggt gtt ggt cga agt tac agg att gcc aaa tct gca gca gca aga   5724
Gly Val Gly Arg Ser Tyr Arg Ile Ala Lys Ser Ala Ala Ala Arg
    1895   1900   1905

aga gcc ctc cga agc ctc aaa gct aat caa cct cag gtt ccc aat   5769
Arg Ala Leu Arg Ser Leu Lys Ala Asn Gln Pro Gln Val Pro Asn
    1910   1915   1920

agc tga   5775
Ser

<210> 2
<211> 1924
<212> PRT
<213> Homo sapiens

<400> 2

Met Lys Ser Pro Ala Leu Gln Pro Leu Ser Met Ala Gly Leu Gln Leu
1   5   10   15

Met Thr Pro Ala Ser Ser Pro Met Gly Pro Phe Phe Gly Leu Pro Trp
            20   25   30

Gln Gln Glu Ala Ile His Asp Asn Ile Tyr Thr Pro Arg Lys Tyr Gln
        35   40   45

Val Glu Leu Leu Glu Ala Ala Leu Asp His Asn Thr Ile Val Cys Leu
        50   55   60

Asn Thr Gly Ser Gly Lys Thr Phe Ile Ala Ser Thr Thr Leu Leu Lys
65   70   75   80

Ser Cys Leu Tyr Leu Asp Leu Gly Glu Thr Ser Ala Arg Asn Gly Lys
          85   90   95

Arg Thr Val Phe Leu Val Asn Ser Ala Asn Gln Val Ala Gln Gln Val
          100   105   110

Ser Ala Val Arg Thr His Ser Asp Leu Lys Val Gly Glu Tyr Ser Asn
          115   120   125

Leu Glu Val Asn Ala Ser Trp Thr Lys Glu Arg Trp Asn Gln Glu Phe
        130   135   140

Thr Lys His Gln Val Leu Ile Met Thr Cys Tyr Val Ala Leu Asn Val
145   150   155   160

Leu Lys Asn Gly Tyr Leu Ser Leu Ser Asp Ile Asn Leu Leu Val Phe
          165   170   175

Asp Glu Cys His Leu Ala Ile Leu Asp His Pro Tyr Arg Glu Phe Met
          180   185   190

Lys Leu Cys Glu Ile Cys Pro Ser Cys Pro Arg Ile Leu Gly Leu Thr
          195   200   205

Ala Ser Ile Leu Asn Gly Lys Trp Asp Pro Glu Asp Leu Glu Glu Lys

210   215   220

Phe Gln Lys Leu Glu Lys Ile Leu Lys Ser Asn Ala Glu Thr Ala Thr
225   230   235   240

Asp Leu Val Val Leu Asp Arg Tyr Thr Ser Gln Pro Cys Glu Ile Val
          245   250   255

Val Asp Cys Gly Pro Phe Thr Asp Arg Ser Gly Leu Tyr Glu Arg Leu
          260   265   270

Leu Met Glu Leu Glu Glu Ala Leu Asn Phe Ile Asn Asp Cys Asn Ile
          275   280   285

Ser Val His Ser Lys Glu Arg Asp Ser Thr Leu Ile Ser Lys Gln Ile
          290   295   300

Leu Ser Asp Cys Arg Ala Val Leu Val Val Leu Gly Pro Trp Cys Ala
305   310   315   320

Asp Lys Val Ala Gly Met Met Val Arg Glu Leu Gln Lys Tyr Ile Lys
          325   330   335

His Glu Gln Glu Glu Leu His Arg Lys Phe Leu Leu Phe Thr Asp Thr
          340   345   350

Phe Leu Arg Lys Ile His Ala Leu Cys Glu Glu His Phe Ser Pro Ala
          355   360   365

Ser Leu Asp Leu Lys Phe Val Thr Pro Lys Val Ile Lys Leu Leu Glu
   370   375   380

Ile Leu Arg Lys Tyr Lys Pro Tyr Glu Arg His Ser Phe Glu Ser Val
385   390   395   400

Glu Trp Tyr Asn Asn Arg Asn Gln Asp Asn Tyr Val Ser Trp Ser Asp
     405   410   415

Ser Glu Asp Asp Asp Glu Asp Glu Glu Ile Glu Glu Lys Glu Lys Pro
     420   425   430

Glu Thr Asn Phe Pro Ser Pro Phe Thr Asn Ile Leu Cys Gly Ile Ile
     435   440   445

Phe Val Glu Arg Arg Tyr Thr Ala Val Val Leu Asn Arg Leu Ile Lys
   450   455   460

Glu Ala Gly Lys Gln Asp Pro Glu Leu Ala Tyr Ile Ser Ser Asn Phe
465   470   475   480

Ile Thr Gly His Gly Ile Gly Lys Asn Gln Pro Arg Asn Asn Thr Met
     485   490   495

Glu Ala Glu Phe Arg Lys Gln Glu Glu Val Leu Arg Lys Phe Arg Ala
     500   505   510

His Glu Thr Asn Leu Leu Ile Ala Thr Ser Ile Val Glu Glu Gly Val
    515   520   525

Asp Ile Pro Lys Cys Asn Leu Val Val Arg Phe Asp Leu Pro Thr Glu
    530   535   540

Tyr Arg Ser Tyr Val Gln Ser Lys Gly Arg Ala Arg Ala Pro Ile Ser
545   550   555   560

Asn Tyr Ile Met Leu Ala Asp Thr Asp Lys Ile Lys Ser Phe Glu Glu
    565   570   575

Asp Leu Lys Thr Tyr Lys Ala Ile Glu Lys Ile Leu Arg Asn Lys Cys
    580   585   590

Ser Lys Ser Val Asp Thr Gly Glu Thr Asp Ile Asp Pro Val Met Asp
    595   600   605

Asp Asp His Val Phe Pro Pro Tyr Val Leu Arg Pro Asp Asp Gly Gly
    610   615   620

Pro Arg Val Thr Ile Asn Thr Ala Ile Gly His Ile Asn Arg Tyr Cys
625   630   635   640

Ala Arg Leu Pro Ser Asp Pro Phe Thr His Leu Ala Pro Lys Cys Arg
    645   650   655

Thr Arg Glu Leu Pro Asp Gly Thr Phe Tyr Ser Thr Leu Tyr Leu Pro
   660   665   670

Ile Asn Ser Pro Leu Arg Ala Ser Ile Val Gly Pro Pro Met Ser Cys
   675   680   685

Val Arg Leu Ala Glu Arg Val Val Ala Leu Ile Cys Cys Glu Lys Leu
   690   695   700

His Lys Ile Gly Glu Leu Asp Asp His Leu Met Pro Val Gly Lys Glu
 705   710   715   720

Thr Val Lys Tyr Glu Glu Glu Leu Asp Leu His Asp Glu Glu Glu Thr
   725   730   735

Ser Val Pro Gly Arg Pro Gly Ser Thr Lys Arg Arg Gln Cys Tyr Pro
   740   745   750

Lys Ala Ile Pro Glu Cys Leu Arg Asp Ser Tyr Pro Arg Pro Asp Gln
   755   760   765

Pro Cys Tyr Leu Tyr Val Ile Gly Met Val Leu Thr Thr Pro Leu Pro
   770   775   780

Asp Glu Leu Asn Phe Arg Arg Arg Lys Leu Tyr Pro Pro Glu Asp Thr
 785   790   795   800

Thr Arg Cys Phe Gly Ile Leu Thr Ala Lys Pro Ile Pro Gln Ile Pro

            805   810   815

His Phe Pro Val Tyr Thr Arg Ser Gly Glu Val Thr Ile Ser Ile Glu
        820   825   830

Leu Lys Lys Ser Gly Phe Met Leu Ser Leu Gln Met Leu Glu Leu Ile
        835   840   845

Thr Arg Leu His Gln Tyr Ile Phe Ser His Ile Leu Arg Leu Glu Lys
        850   855   860

Pro Ala Leu Glu Phe Lys Pro Thr Asp Ala Asp Ser Ala Tyr Cys Val
865   870   875   880

Leu Pro Leu Asn Val Val Asn Asp Ser Ser Thr Leu Asp Ile Asp Phe
            885   890   895

Lys Phe Met Glu Asp Ile Glu Lys Ser Glu Ala Arg Ile Gly Ile Pro
        900   905   910

Ser Thr Lys Tyr Thr Lys Glu Thr Pro Phe Val Phe Lys Leu Glu Asp
        915   920   925

Tyr Gln Asp Ala Val Ile Ile Pro Arg Tyr Arg Asn Phe Asp Gln Pro
        930   935   940

His Arg Phe Tyr Val Ala Asp Val Tyr Thr Asp Leu Thr Pro Leu Ser
945   950   955   960

Lys Phe Pro Ser Pro Glu Tyr Glu Thr Phe Ala Glu Tyr Tyr Lys Thr
        965   970   975

Lys Tyr Asn Leu Asp Leu Thr Asn Leu Asn Gln Pro Leu Leu Asp Val
        980   985   990

Asp His Thr Ser Ser Arg Leu Asn Leu Leu Thr Pro Arg His Leu Asn
        995   1000   1005

Gln Lys Gly Lys Ala Leu Pro Leu Ser Ser Ala Glu Lys Arg Lys
       1010   1015   1020

Ala Lys Trp Glu Ser Leu Gln Asn Lys Gln Ile Leu Val Pro Glu
       1025   1030   1035

Leu Cys Ala Ile His Pro Ile Pro Ala Ser Leu Trp Arg Lys Ala
       1040   1045   1050

Val Cys Leu Pro Ser Ile Leu Tyr Arg Leu His Cys Leu Leu Thr
       1055   1060   1065

Ala Glu Glu Leu Arg Ala Gln Thr Ala Ser Asp Ala Gly Val Gly
       1070   1075   1080

Val Arg Ser Leu Pro Ala Asp Phe Arg Tyr Pro Asn Leu Asp Phe
       1085   1090   1095

Gly Trp Lys Lys Ser Ile Asp Ser Lys Ser Phe Ile Ser Ile Ser
   1100   1105   1110

Asn Ser Ser Ser Ala Glu Asn Asp Asn Tyr Cys Lys His Ser Thr
   1115   1120   1125

Ile Val Pro Glu Asn Ala Ala His Gln Gly Ala Asn Arg Thr Ser
   1130   1135   1140

Ser Leu Glu Asn His Asp Gln Met Ser Val Asn Cys Arg Thr Leu
   1145   1150   1155

Leu Ser Glu Ser Pro Gly Lys Leu His Val Glu Val Ser Ala Asp
   1160   1165   1170

Leu Thr Ala Ile Asn Gly Leu Ser Tyr Asn Gln Asn Leu Ala Asn
   1175   1180   1185

Gly Ser Tyr Asp Leu Ala Asn Arg Asp Phe Cys Gln Gly Asn Gln
   1190   1195   1200

Leu Asn Tyr Tyr Lys Gln Glu Ile Pro Val Gln Pro Thr Thr Ser
   1205   1210   1215

Tyr Ser Ile Gln Asn Leu Tyr Ser Tyr Glu Asn Gln Pro Gln Pro
   1220   1225   1230

Ser Asp Glu Cys Thr Leu Leu Ser Asn Lys Tyr Leu Asp Gly Asn
   1235   1240   1245

Ala Asn Lys Ser Thr Ser Asp Gly Ser Pro Val Met Ala Val Met
   1250   1255   1260

Pro Gly Thr Thr Asp Thr Ile Gln Val Leu Lys Gly Arg Met Asp
   1265   1270   1275

Ser Glu Gln Ser Pro Ser Ile Gly Tyr Ser Ser Arg Thr Leu Gly
   1280   1285   1290

Pro Asn Pro Gly Leu Ile Leu Gln Ala Leu Thr Leu Ser Asn Ala
   1295   1300   1305

Ser Asp Gly Phe Asn Leu Glu Arg Leu Glu Met Leu Gly Asp Ser
   1310   1315   1320

Phe Leu Lys His Ala Ile Thr Thr Tyr Leu Phe Cys Thr Tyr Pro
   1325   1330   1335

Asp Ala His Glu Gly Arg Leu Ser Tyr Met Arg Ser Lys Lys Val
   1340   1345   1350

Ser Asn Cys Asn Leu Tyr Arg Leu Gly Lys Lys Lys Gly Leu Pro
   1355   1360   1365

Ser Arg Met Val Val Ser Ile Phe Asp Pro Pro Val Asn Trp Leu

```
          1370            1375            1380


Pro Pro  Gly Tyr Val Val Asn  Gln Asp Lys Ser Asn  Thr Asp Lys
     1385            1390            1395


Trp Glu  Lys Asp Glu Met Thr  Lys Asp Cys Met Leu  Ala Asn Gly
     1400            1405            1410


Lys Leu  Asp Glu Asp Tyr Glu  Glu Glu Asp Glu Glu  Glu Glu Ser
     1415            1420            1425


Leu Met  Trp Arg Ala Pro Lys  Glu Glu Ala Asp Tyr  Glu Asp Asp
     1430            1435            1440


Phe Leu  Glu Tyr Asp Gln Glu  His Ile Arg Phe Ile  Asp Asn Met
     1445            1450            1455


Leu Met  Gly Ser Gly Ala Phe  Val Lys Lys Ile Ser  Leu Ser Pro
     1460            1465            1470


Phe Ser  Thr Thr Asp Ser Ala  Tyr Glu Trp Lys Met  Pro Lys Lys
     1475            1480            1485


Ser Ser  Leu Gly Ser Met Pro  Phe Ser Ser Asp Phe  Glu Asp Phe
     1490            1495            1500


Asp Tyr  Ser Ser Trp Asp Ala  Met Cys Tyr Leu Asp  Pro Ser Lys
     1505            1510            1515
```

Ala Val Glu Glu Asp Asp Phe Val Val Gly Phe Trp Asn Pro Ser
    1520   1525   1530

Glu Glu Asn Cys Gly Val Asp Thr Gly Lys Gln Ser Ile Ser Tyr
    1535   1540   1545

Asp Leu His Thr Glu Gln Cys Ile Ala Asp Lys Ser Ile Ala Asp
    1550   1555   1560

Cys Val Glu Ala Leu Leu Gly Cys Tyr Leu Thr Ser Cys Gly Glu
    1565   1570   1575

Arg Ala Ala Gln Leu Phe Leu Cys Ser Leu Gly Leu Lys Val Leu
    1580   1585   1590

Pro Val Ile Lys Arg Thr Asp Arg Glu Lys Ala Leu Cys Pro Thr
    1595   1600   1605

Arg Glu Asn Phe Asn Ser Gln Gln Lys Asn Leu Ser Val Ser Cys
    1610   1615   1620

Ala Ala Ala Ser Val Ala Ser Ser Arg Ser Ser Val Leu Lys Asp
    1625   1630   1635

Ser Glu Tyr Gly Cys Leu Lys Ile Pro Pro Arg Cys Met Phe Asp
    1640   1645   1650

His Pro   Asp Ala Asp Lys Thr   Leu Asn His Leu Ile   Ser Gly Phe
   1655   1660   1665

Glu Asn   Phe Glu Lys Lys Ile   Asn Tyr Arg Phe Lys   Asn Lys Ala
   1670   1675   1680

Tyr Leu   Leu Gln Ala Phe Thr   His Ala Ser Tyr His   Tyr Asn Thr
   1685   1690   1695

Ile Thr   Asp Cys Tyr Gln Arg   Leu Glu Phe Leu Gly   Asp Ala Ile
   1700   1705   1710

Leu Asp   Tyr Leu Ile Thr Lys   His Leu Tyr Glu Asp   Pro Arg Gln
   1715   1720   1725

His Ser   Pro Gly Val Leu Thr   Asp Leu Arg Ser Ala   Leu Val Asn
   1730   1735   1740

Asn Thr   Ile Phe Ala Ser Leu   Ala Val Lys Tyr Asp   Tyr His Lys
   1745   1750   1755

Tyr Phe   Lys Ala Val Ser Pro   Glu Leu Phe His Val   Ile Asp Asp
   1760   1765   1770

Phe Val   Gln Phe Gln Leu Glu   Lys Asn Glu Met Gln   Gly Met Asp
   1775   1780   1785

Ser Glu Leu Arg Arg Ser Glu Glu Asp Glu Glu Lys Glu Glu Asp
    1790   1795   1800

Ile Glu Val Pro Lys Ala Met Gly Asp Ile Phe Glu Ser Leu Ala
    1805   1810   1815

Gly Ala Ile Tyr Met Asp Ser Gly Met Ser Leu Glu Thr Val Trp
    1820   1825   1830

Gln Val Tyr Tyr Pro Met Met Arg Pro Leu Ile Glu Lys Phe Ser
    1835   1840   1845

Ala Asn Val Pro Arg Ser Pro Val Arg Glu Leu Leu Glu Met Glu
    1850   1855   1860

Pro Glu Thr Ala Lys Phe Ser Pro Ala Glu Arg Thr Tyr Asp Gly
    1865   1870   1875

Lys Val Arg Val Thr Val Glu Val Val Gly Lys Gly Lys Phe Lys
    1880   1885   1890

Gly Val Gly Arg Ser Tyr Arg Ile Ala Lys Ser Ala Ala Ala Arg
    1895   1900   1905

Arg Ala Leu Arg Ser Leu Lys Ala Asn Gln Pro Gln Val Pro Asn
    1910   1915   1920

Ser

<210> 3
<211> 6750
<212> DNA
<213> Drosophila melanogaster

<220>
<221> CDS
<222> (1)..(6750)

<400> 3
atg gcg ttc cac tgg tgc gac aac aat ctg cac acc acc gtg ttc acg   48
Met Ala Phe His Trp Cys Asp Asn Asn Leu His Thr Thr Val Phe Thr
1   5   10   15

ccg cgc gac ttt cag gtg gag cta ctg gcc acc gcc tac gag cgg aac   96
Pro Arg Asp Phe Gln Val Glu Leu Leu Ala Thr Ala Tyr Glu Arg Asn
                20   25   30

acg att att tgc ctg ggc cat cga agt tcc aag gag ttt ata gcc ctc   144
Thr Ile Ile Cys Leu Gly His Arg Ser Ser Lys Glu Phe Ile Ala Leu
            35   40   45

aag ctg ctc cag gag ctg tcg cgt cga gca cgc cga cat ggt cgt gtc   192
Lys Leu Leu Gln Glu Leu Ser Arg Arg Ala Arg Arg His Gly Arg Val
        50   55   60

agt gtc tat ctc agt tgc gag gtt ggc acc agc acg gaa cca tgc tcc   240
Ser Val Tyr Leu Ser Cys Glu Val Gly Thr Ser Thr Glu Pro Cys Ser
65   70   75   80

atc tac acg atg ctc acc cac ttg act gac ctg cgg gtg tgg cag gag   288
Ile Tyr Thr Met Leu Thr His Leu Thr Asp Leu Arg Val Trp Gln Glu
                85   90   95

29

cag ccg gat atg caa att ccc ttt gat cat tgc tgg acg gac tat cac   336
Gln Pro Asp Met Gln Ile Pro Phe Asp His Cys Trp Thr Asp Tyr His
        100   105   110

gtt tcc atc cta cgg cca gag gga ttt ctt tat ctg ctc gaa act cgc   384
Val Ser Ile Leu Arg Pro Glu Gly Phe Leu Tyr Leu Leu Glu Thr Arg
        115   120   125

gag ctg ctg ctg agc agc gtc gaa ctg atc gtg ctg gaa gat tgt cat   432
Glu Leu Leu Leu Ser Ser Val Glu Leu Ile Val Leu Glu Asp Cys His
        130   135   140

gac agc gcc gtt tat cag agg ata agg cct ctg ttc gag aat cac att   480
Asp Ser Ala Val Tyr Gln Arg Ile Arg Pro Leu Phe Glu Asn His Ile
145   150   155   160

atg cca gcg cca ccg gcg gac agg cca cgg att ctc gga ctc gct gga   528
Met Pro Ala Pro Pro Ala Asp Arg Pro Arg Ile Leu Gly Leu Ala Gly
        165   170   175

ccg ctg cac agc gcc gga tgt gag ctg cag caa ctg agc gcc atg ctg   576
Pro Leu His Ser Ala Gly Cys Glu Leu Gln Gln Leu Ser Ala Met Leu
        180   185   190

gcc acc ctg gag cag agt gtg ctt tgc cag atc gag acg gcc agt gat   624
Ala Thr Leu Glu Gln Ser Val Leu Cys Gln Ile Glu Thr Ala Ser Asp
        195   200   205

att gtc acc gtg ttg cgt tac tgt tcc cga ccg cac gaa tac atc gta   672
Ile Val Thr Val Leu Arg Tyr Cys Ser Arg Pro His Glu Tyr Ile Val
        210   215   220

cag tgc gcc ccc ttc gag atg gac gaa ctg tcc ctg gtg ctt gcc gat   720
Gln Cys Ala Pro Phe Glu Met Asp Glu Leu Ser Leu Val Leu Ala Asp
225 .   230   235   240

gtg ctc aac aca cac aag tcc ttt tta ttg gac cac cgc tac gat ccc   768

Val Leu Asn Thr His Lys Ser Phe Leu Leu Asp His Arg Tyr Asp Pro
          245   250   255

tac gaa atc tac ggc aca gac cag ttt atg gac gaa ctg aaa gac ata   816
Tyr Glu Ile Tyr Gly Thr Asp Gln Phe Met Asp Glu Leu Lys Asp Ile
          260   265   270

ccc gat ccc aag gtg gac ccc ctg aac gtc atc aac tca cta ctg gtc   864
Pro Asp Pro Lys Val Asp Pro Leu Asn Val Ile Asn Ser Leu Leu Val
           275   280   285

gtg ctg cac gag atg ggt cct tgg tgc acg cag cgg gct gca cat cac   912
Val Leu His Glu Met Gly Pro Trp Cys Thr Gln Arg Ala Ala His His
        290   295   300

ttt tac caa tgc aat gag aag tta aag gtg aag acg ccg cac gaa cgt   960
Phe Tyr Gln Cys Asn Glu Lys Leu Lys Val Lys Thr Pro His Glu Arg
305   310   315   320

cac tac ttg ctg tac tgc cta gtg agc acg gcc ctt atc caa ctg tac   1008
His Tyr Leu Leu Tyr Cys Leu Val Ser Thr Ala Leu Ile Gln Leu Tyr
         325   330   335

tcc ctc tgc gaa cac gca ttc cat cga cat tta gga agt ggc agc gat   1056
Ser Leu Cys Glu His Ala Phe His Arg His Leu Gly Ser Gly Ser Asp
        340   345   350

tca cgc cag acc atc gaa cgc tat tcc agc ccc aag gtg cga cgt ctg   1104
Ser Arg Gln Thr Ile Glu Arg Tyr Ser Ser Pro Lys Val Arg Arg Leu
        355   360   365

ttg cag aca ctg agg tgc ttc aag ccg gaa gag gtg cac acc caa gcg   1152
Leu Gln Thr Leu Arg Cys Phe Lys Pro Glu Glu Val His Thr Gln Ala
        370   375   380

gac gga ctg cgc aga atg cgg cat cag gtg gat cag gcg gac ttc aat   1200
Asp Gly Leu Arg Arg Met Arg His Gln Val Asp Gln Ala Asp Phe Asn

385   390   395   400

```
cgg tta tct cat acg ctg gaa agc aag tgc cga atg gtg gat caa atg    1248
Arg Leu Ser His Thr Leu Glu Ser Lys Cys Arg Met Val Asp Gln Met
              405            410            415


gac caa ccg ccg acg gag aca cga gcc ctg gtg gcc act ctt gag cag    1296
Asp Gln Pro Pro Thr Glu Thr Arg Ala Leu Val Ala Thr Leu Glu Gln
         420            425            430


att ctg cac acg aca gag gac agg cag acg aac aga agc gcc gct cgg    1344
Ile Leu His Thr Thr Glu Asp Arg Gln Thr Asn Arg Ser Ala Ala Arg
       435            440            445


gtg act cct act cct act ccc gct cat gcg aag ccg aaa cct agc tct    1392
Val Thr Pro Thr Pro Thr Pro Ala His Ala Lys Pro Lys Pro Ser Ser
     450            455            460


ggt gcc aac act gca caa cca cga act cgt aga cgt gtg tac acc agg    1440
Gly Ala Asn Thr Ala Gln Pro Arg Thr Arg Arg Arg Val Tyr Thr Arg .
465            470            475            480


cgc cac cac cgg gat cac aat gat ggc agc gac acg ctc tgc gca ctg    1488
Arg His His Arg Asp His Asn Asp Gly Ser Asp Thr Leu Cys Ala Leu
              485            490            495


att tac tgc aac cag aac cac acg gct cgc gtg ctc ttt gag ctt cta    1536
Ile Tyr Cys Asn Gln Asn His Thr Ala Arg Val Leu Phe Glu Leu Leu
         500            505            510


gcg gag att agc aga cgt gat ccc gat ctc aag ttc cta cgc tgc cag    1584
Ala Glu Ile Ser Arg Arg Asp Pro Asp Leu Lys Phe Leu Arg Cys Gln
       515            520            525


tac acc acg gac cgg gtg gca gat ccc acc acg gag ccc aaa gag gct    1632
Tyr Thr Thr Asp Arg Val Ala Asp Pro Thr Thr Glu Pro Lys Glu Ala
     530            535            540
```

gag ttg gag cac cgg cgg cag gaa gag gtg cta aag cgc ttc cgc atg   1680
Glu Leu Glu His Arg Arg Gln Glu Glu Val Leu Lys Arg Phe Arg Met
545   550   555   560

cat gac tgc aat gtc ctg atc ggt act tcg gtg ctg gaa gag ggc atc   1728
His Asp Cys Asn Val Leu Ile Gly Thr Ser Val Leu Glu Glu Gly Ile
            565   570   575

gat gtg ccc aag tgc aat ttg gtt gtg cgc tgg gat ccg cca acc aca   1776
Asp Val Pro Lys Cys Asn Leu Val Val Arg Trp Asp Pro Pro Thr Thr
            580   585   590

tat cgc agt tac gtt cag tgc aaa ggt cga gcc cgt gct gct cca gcc   1824
Tyr Arg Ser Tyr Val Gln Cys Lys Gly Arg Ala Arg Ala Ala Pro Ala
            595   600   605

tat cat gtc att ctc gtc gct ccg agt tat aaa agc cca act gtg ggg   1872
Tyr His Val Ile Leu Val Ala Pro Ser Tyr Lys Ser Pro Thr Val Gly
            610   615   620

tca gtg cag ctg acc gat cgg agt cat cgg tat att tgc gcg act ggt   1920
Ser Val Gln Leu Thr Asp Arg Ser His Arg Tyr Ile Cys Ala Thr Gly
625   630   635   640

gat act aca gag gcg gac agc gac tct gat gat tca gcg atg cca aac   1968
Asp Thr Thr Glu Ala Asp Ser Asp Ser Asp Asp Ser Ala Met Pro Asn
            645   650   655

tcg tcc ggc tcg gat ccc tat act ttt ggc acg gca cgc gga acc gtg   2016
Ser Ser Gly Ser Asp Pro Tyr Thr Phe Gly Thr Ala Arg Gly Thr Val
            660   665   670

aag atc ctc aac ccc gaa gtg ttc agt aaa caa cca ccg aca gcg tgc   2064
Lys Ile Leu Asn Pro Glu Val Phe Ser Lys Gln Pro Pro Thr Ala Cys
            675   680   685

gac att aag ctg cag gag atc cag gac gaa ttg cca gcc gca gcg cag   2112
Asp Ile Lys Leu Gln Glu Ile Gln Asp Glu Leu Pro Ala Ala Ala Gln
     690   695   700

ctg gat acg agc aac tcc agc gac gaa gcc gtc agc atg agt aac acg   2160
Leu Asp Thr Ser Asn Ser Ser Asp Glu Ala Val Ser Met Ser Asn Thr
705   710   715   720

tct cca agc gag agc agt aca gaa caa aaa tcc aga cgc ttc cag tgc   2208
Ser Pro Ser Glu Ser Ser Thr Glu Gln Lys Ser Arg Arg Phe Gln Cys
        725   730   735

gag ctg agc tct tta acg gag cca gaa gac aca agt gat act aca gcc   2256
Glu Leu Ser Ser Leu Thr Glu Pro Glu Asp Thr Ser Asp Thr Thr Ala
        740   745   750

gaa atc gat act gct cat agt tta gcc agc acc acg aag gac ttg gtg   2304
Glu Ile Asp Thr Ala His Ser Leu Ala Ser Thr Thr Lys Asp Leu Val
        755   760   765

cat caa atg gca cag tat cgc gaa atc gag cag atg ctg cta tcc aag   2352
His Gln Met Ala Gln Tyr Arg Glu Ile Glu Gln Met Leu Leu Ser Lys
     770   775   780

tgc gcc aac aca gag ccg ccg gag cag gag cag agt gag gcg gaa cgt   2400
Cys Ala Asn Thr Glu Pro Pro Glu Gln Glu Gln Ser Glu Ala Glu Arg
785   790   795   800

ttt agt gcc tgc ctg gcc gca tac cga ccc aag ccg cac ctg cta aca   2448
Phe Ser Ala Cys Leu Ala Ala Tyr Arg Pro Lys Pro His Leu Leu Thr
        805   810   815

ggc gcc tcc gtg gat ctg ggt tct gct ata gct ttg gtc aac aag tac   2496
Gly Ala Ser Val Asp Leu Gly Ser Ala Ile Ala Leu Val Asn Lys Tyr
        820   825   830

tgc gcc cga ctg cca agc gac acg ttc acc aag ttg acg gcg ttg tgg   2544

Cys Ala Arg Leu Pro Ser Asp Thr Phe Thr Lys Leu Thr Ala Leu Trp
    835   840   845

cgc tgc acc cga aac gaa agg gct gga gtg acc ctg ttt cag tac aca   2592
Arg Cys Thr Arg Asn Glu Arg Ala Gly Val Thr Leu Phe Gln Tyr Thr
    850   855   860

ctc cgt ctg ccc atc aac tcg cca ttg aag cat gac att gtg ggt ctt   2640
Leu Arg Leu Pro Ile Asn Ser Pro Leu Lys His Asp Ile Val Gly Leu
    865   870   875   880

ccg atg cca act caa aca ttg gcc cgc cga ctg gct gcc ttg cag gct   2688
Pro Met Pro Thr Gln Thr Leu Ala Arg Arg Leu Ala Ala Leu Gln Ala
        885   890   895

tgc gtg gaa ctg cac agg atc ggt gag tta gac gat cag ttg cag cct   2736
Cys Val Glu Leu His Arg Ile Gly Glu Leu Asp Asp Gln Leu Gln Pro
      900   905   910

atc ggc aag gag gga ttt cgt gcc ctg gag ccg gac tgg gag tgc ttt   2784
Ile Gly Lys Glu Gly Phe Arg Ala Leu Glu Pro Asp Trp Glu Cys Phe
      915   920   925

gaa ctg gag cca gag gac gaa cag att gtg cag cta agc gat gaa cca   2832
Glu Leu Glu Pro Glu Asp Glu Gln Ile Val Gln Leu Ser Asp Glu Pro
    930   935   940

cgt ccg gga aca acg aag cgt cgt cag tac tat tac aaa cgc att gca   2880
Arg Pro Gly Thr Thr Lys Arg Arg Gln Tyr Tyr Tyr Lys Arg Ile Ala
    945   950   955   960

tcc gaa ttt tgc gat tgc cgt ccc gtt gcc gga gcg cca tgc tat ttg   2928
Ser Glu Phe Cys Asp Cys Arg Pro Val Ala Gly Ala Pro Cys Tyr Leu
      965   970   975

tac ttt atc caa ctg acg ctc caa tgt ccg att ccc gaa gag caa aac   2976
Tyr Phe Ile Gln Leu Thr Leu Gln Cys Pro Ile Pro Glu Glu Gln Asn

              980   985   990

acg cgg gga cgc aag att tat ccg ccc gaa gat gcg cag cag gga ttc   3024
Thr Arg Gly Arg Lys Ile Tyr Pro Pro Glu Asp Ala Gln Gln Gly Phe
       995   1000   1005

ggc att cta acc acc aaa cgg ata ccc aag ctg agt gct ttc tcg   3069
Gly Ile Leu Thr Thr Lys Arg Ile Pro Lys Leu Ser Ala Phe Ser
       1010   1015   1020

ata ttc acg cgt tcc ggt gag gtg aag gtt tcc ctg gag tta gct   3114
Ile Phe Thr Arg Ser Gly Glu Val Lys Val Ser Leu Glu Leu Ala
       1025   1030   1035

aag gaa cgc gtg att cta act agc gaa caa ata gtc tgc atc aac   3159
Lys Glu Arg Val Ile Leu Thr Ser Glu Gln Ile Val Cys Ile Asn
       1040   1045   1050

gga ttt tta aac tac acg ttc acc aat gta ctg cgt ttg caa aag   3204
Gly Phe Leu Asn Tyr Thr Phe Thr Asn Val Leu Arg Leu Gln Lys
       1055   1060   1065

ttt ctg atg ctc ttc gat ccg gac tcc acg gaa aat tgt gta ttc   3249
Phe Leu Met Leu Phe Asp Pro Asp Ser Thr Glu Asn Cys Val Phe
       1070   1075   1080

att gtg ccc acc gtg aag gca cca gct ggc ggc aag cac atc gac   3294
Ile Val Pro Thr Val Lys Ala Pro Ala Gly Gly Lys His Ile Asp
       1085   1090   1095

tgg cag ttt ctg gag ctg atc caa gcg aat gga aat aca atg cca   3339
Trp Gln Phe Leu Glu Leu Ile Gln Ala Asn Gly Asn Thr Met Pro
       1100   1105   1110

cgg gca gtg ccc gat gag gag cgc cag gcg cag ccg ttt gat ccg   3384
Arg Ala Val Pro Asp Glu Glu Arg Gln Ala Gln Pro Phe Asp Pro
       1115   1120   1125

caa cgc ttc cag gat gcc gtc gtt atg ccg tgg tat cgc aac cag   3429
Gln Arg Phe Gln Asp Ala Val Val Met Pro Trp Tyr Arg Asn Gln
   1130   1135   1140

gat caa ccg cag tat ttc tat gtg gcg gag ata tgt cca cat cta   3474
Asp Gln Pro Gln Tyr Phe Tyr Val Ala Glu Ile Cys Pro His Leu
   1145   1150   1155

tcc cca ctc agc tgc ttc cct ggt gac aac tac cgc acg ttc aag   3519
Ser Pro Leu Ser Cys Phe Pro Gly Asp Asn Tyr Arg Thr Phe Lys
   1160   1165   1170

cac tac tac ctc gtc aag tat ggt ctg acc ata cag aat acc tcg   3564
His Tyr Tyr Leu Val Lys Tyr Gly Leu Thr Ile Gln Asn Thr Ser
   1175   1180   1185

cag ccg cta ttg gac gtg gat cac acc agt gcg cgg tta aac ttc   3609
Gln Pro Leu Leu Asp Val Asp His Thr Ser Ala Arg Leu Asn Phe
   1190   1195   1200

ctc acg cca cga tac gtt aat cgc aag ggc gtt gct ctg ccc act   3654
Leu Thr Pro Arg Tyr Val Asn Arg Lys Gly Val Ala Leu Pro Thr
   1205   1210   1215

agt tcg gag gag aca aag cgg gca aag cgc gag aat ctc gaa cag   3699
Ser Ser Glu Glu Thr Lys Arg Ala Lys Arg Glu Asn Leu Glu Gln
   1220   1225   1230

aag cag atc ctt gtg cca gag ctc tgc act gtg cat cca ttc ccc   3744
Lys Gln Ile Leu Val Pro Glu Leu Cys Thr Val His Pro Phe Pro
   1235   1240   1245

gcc tcc ttg tgg cga act gcc gtg tgc ctg ccc tgc atc ctg tac   3789
Ala Ser Leu Trp Arg Thr Ala Val Cys Leu Pro Cys Ile Leu Tyr
   1250   1255   1260

cgc ata aat ggt ctt cta ttg gcc gac gat att cgg aaa cag gtt   3834
Arg Ile Asn Gly Leu Leu Leu Ala Asp Asp Ile Arg Lys Gln Val
    1265   1270   1275

tct gcg gat ctg ggg ctg gga agg caa cag atc gaa gat gag gat   3879
Ser Ala Asp Leu Gly Leu Gly Arg Gln Gln Ile Glu Asp Glu Asp
    1280   1285   1290

ttc gag tgg ccc atg ctg gac ttt ggg tgg agt cta tcg gag gtg   3924
Phe Glu Trp Pro Met Leu Asp Phe Gly Trp Ser Leu Ser Glu Val
    1295   1300   1305

ctc aag aaa tcg cgg gag tcc aaa caa aag gag tcc ctt aag gat   3969
Leu Lys Lys Ser Arg Glu Ser Lys Gln Lys Glu Ser Leu Lys Asp
    1310   1315   1320

gat act att aat ggc aaa gac tta gct gat gtt gaa aag aaa ccg   4014
Asp Thr Ile Asn Gly Lys Asp Leu Ala Asp Val Glu Lys Lys Pro
    1325   1330   1335

act agc gag gag acc caa cta gat aag gat tca aaa gac gat aag   4059
Thr Ser Glu Glu Thr Gln Leu Asp Lys Asp Ser Lys Asp Asp Lys
    1340   1345   1350

gtt gag aaa agt gct att gaa cta atc att gag gga gag gag aag   4104
Val Glu Lys Ser Ala Ile Glu Leu Ile Ile Glu Gly Glu Glu Lys
    1355   1360   1365

ctg caa gag gct gat gac ttc att gag ata ggc act tgg tca aac   4149
Leu Gln Glu Ala Asp Asp Phe Ile Glu Ile Gly Thr Trp Ser Asn
    1370   1375   1380

gat atg gcc gac gat ata gct agt ttt aac caa gaa gac gac gac   4194
Asp Met Ala Asp Asp Ile Ala Ser Phe Asn Gln Glu Asp Asp Asp
    1385   1390   1395

gag gat gac gcc ttc cat ctc cca gtt tta ccg gca aac gtt aag   4239

Glu Asp Asp Ala Phe His Leu Pro Val Leu Pro Ala Asn Val Lys
   1400   1405   1410

ttc tgt gat cag caa acg cgc tac ggt tcg ccc aca ttt tgg gat   4284
Phe Cys Asp Gln Gln Thr Arg Tyr Gly Ser Pro Thr Phe Trp Asp
   1415   1420   1425

gtg agc aat ggc gaa agc ggc ttc aag ggt cca aag agc agt cag   4329
Val Ser Asn Gly Glu Ser Gly Phe Lys Gly Pro Lys Ser Ser Gln
   1430   1435   1440

aat aag cag ggt ggc aag ggc aaa gca aag ggt ccg gca aag ccc   4374
Asn Lys Gln Gly Gly Lys Gly Lys Ala Lys Gly Pro Ala Lys Pro
   1445   1450   1455

aca ttt aac tat tat gac tcg gac aat tcg ctg ggt tcc agc tac   4419
Thr Phe Asn Tyr Tyr Asp Ser Asp Asn Ser Leu Gly Ser Ser Tyr
   1460   1465   1470

gat gac gac gat aac gca ggt ccg ctc aat tac atg cat cac aac   4464
Asp Asp Asp Asp Asn Ala Gly Pro Leu Asn Tyr Met His His Asn
   1475   1480   1485

tac agt tcg gat gac gac gat gtg gca gat gat atc gat gcg gga   4509
Tyr Ser Ser Asp Asp Asp Asp Val Ala Asp Asp Ile Asp Ala Gly
   1490   1495   1500

cgc att gcg ttc acc tcc aag aat gaa gcg gag act att gaa acc   4554
Arg Ile Ala Phe Thr Ser Lys Asn Glu Ala Glu Thr Ile Glu Thr
   1505   1510   1515

gca cag gaa gtg gaa aag cgc cag aag cag ctg tcc atc atc cag   4599
Ala Gln Glu Val Glu Lys Arg Gln Lys Gln Leu Ser Ile Ile Gln
   1520   1525   1530

gcg acc aat gct aac gag cgg cag tat cag cag aca aag aac ctg   4644
Ala Thr Asn Ala Asn Glu Arg Gln Tyr Gln Gln Thr Lys Asn Leu

```
            1535              1540              1545


ctc att  gga ttc aat ttt aag  cat gag gac cag aag  gaa cct gcc    4689
Leu Ile  Gly Phe Asn Phe Lys  His Glu Asp Gln Lys  Glu Pro Ala
     1550             1555              1560


act ata  aga tat gaa gaa tcc  ata gct aag ctc aaa  acg gaa ata    4734
Thr Ile  Arg Tyr Glu Glu Ser  Ile Ala Lys Leu Lys  Thr Glu Ile
     1565             1570              1575


gaa tcc  ggc ggc atg ttg gtg  ccg cac gac cag cag  ttg gtt cta    4779
Glu Ser  Gly Gly Met Leu Val  Pro His Asp Gln Gln  Leu Val Leu
     1580             1585              1590


aaa aga  agt gat gcc gct gag  gct cag gtt gca aag  gta tcg atg    4824
Lys Arg  Ser Asp Ala Ala Glu  Ala Gln Val Ala Lys  Val Ser Met
     1595             1600              1605


atg gag  cta ttg aag cag ctg  ctg ccg tat gta aat  gaa gat gtg    4869
Met Glu  Leu Leu Lys Gln Leu  Leu Pro Tyr Val Asn  Glu Asp Val
     1610             1615              1620


ctg gcc  aaa aag ctg ggt gat  agg cgc gag ctt ctg  ctg tcg gat    4914
Leu Ala  Lys Lys Leu Gly Asp  Arg Arg Glu Leu Leu  Leu Ser Asp
     1625             1630              1635


ttg gta  gag cta aat gca gat  tgg gta gcg cga cat  gag cag gag    4959
Leu Val  Glu Leu Asn Ala Asp  Trp Val Ala Arg His  Glu Gln Glu
     1640             1645              1650


acc tac  aat gta atg gga tgc  gga gat agt ttt gac  aac tat aac    5004
Thr Tyr  Asn Val Met Gly Cys  Gly Asp Ser Phe Asp  Asn Tyr Asn
     1655             1660              1665


gat cat  cat cgg ctg aac ttg  gat gaa aag caa ctg  aaa ctg caa    5049
Asp His  His Arg Leu Asn Leu  Asp Glu Lys Gln Leu  Lys Leu Gln
     1670             1675              1680
```

tac gaa cga att gaa att gag cca cct act tcc acg aag gcc ata   5094
Tyr Glu Arg Ile Glu Ile Glu Pro Pro Thr Ser Thr Lys Ala Ile
     1685   1690   1695

acc tca gcc ata tta cca gct ggc ttc agt ttc gat cga caa ccg   5139
Thr Ser Ala Ile Leu Pro Ala Gly Phe Ser Phe Asp Arg Gln Pro
     1700   1705   1710

gat cta gtg ggc cat cca gga ccc agt ccc agc atc att ttg caa   5184
Asp Leu Val Gly His Pro Gly Pro Ser Pro Ser Ile Ile Leu Gln
     1715   1720   1725

gcc ctc aca atg tcc aat gct aac gat ggc atc aat ctg gag cga   5229
Ala Leu Thr Met Ser Asn Ala Asn Asp Gly Ile Asn Leu Glu Arg
     1730   1735   1740

ctg gag aca att gga gat tcc ttt cta aag tat gcc att acc acc   5274
Leu Glu Thr Ile Gly Asp Ser Phe Leu Lys Tyr Ala Ile Thr Thr
     1745   1750   1755

tac ttg tac atc acc tac gag aat gtg cac gag gga aaa cta agt   5319
Tyr Leu Tyr Ile Thr Tyr Glu Asn Val His Glu Gly Lys Leu Ser
     1760   1765   1770

cac ctg cgc tcc aag cag gtt gcc aat ctc aat ctc tat cgt ctg   5364
His Leu Arg Ser Lys Gln Val Ala Asn Leu Asn Leu Tyr Arg Leu
     1775   1780   1785

ggc aga cgt aag aga ctg ggt gaa tat atg ata gcc act aaa ttc   5409
Gly Arg Arg Lys Arg Leu Gly Glu Tyr Met Ile Ala Thr Lys Phe
     1790   1795   1800

gag ccg cac gac aat tgg ctg cca ccc tgc tac tac gtg cca aag   5454
Glu Pro His Asp Asn Trp Leu Pro Pro Cys Tyr Tyr Val Pro Lys
     1805   1810   1815

gag cta gag aag gcg ctc atc gag gcg aag atc ccc act cac cat   5499
Glu Leu Glu Lys Ala Leu Ile Glu Ala Lys Ile Pro Thr His His
    1820   1825   1830

tgg aag ctg gcc gat ctg cta gac att aag aac cta agc agt gtg   5544
Trp Lys Leu Ala Asp Leu Leu Asp Ile Lys Asn Leu Ser Ser Val
    1835   1840   1845

caa atc tgc gag atg gtt cgc gaa aaa gcc gat gcc ctg ggc ttg   5589
Gln Ile Cys Glu Met Val Arg Glu Lys Ala Asp Ala Leu Gly Leu
    1850   1855   1860

gag cag aat ggg ggt gcc caa aat gga caa ctt gac gac tcc aat   5634
Glu Gln Asn Gly Gly Ala Gln Asn Gly Gln Leu Asp Asp Ser Asn
    1865   1870   1875

gat agc tgc aat gat ttt agc tgt ttt att ccc tac aac ctt gtt   5679
Asp Ser Cys Asn Asp Phe Ser Cys Phe Ile Pro Tyr Asn Leu Val
    1880   1885   1890

tcg caa cac agc att ccg gat aag tct att gcc gat tgc gtc gaa   5724
Ser Gln His Ser Ile Pro Asp Lys Ser Ile Ala Asp Cys Val Glu
    1895   1900   1905

gcc ctc att gga gcc tat ctc att gag tgc gga ccc cga ggg gct   5769
Ala Leu Ile Gly Ala Tyr Leu Ile Glu Cys Gly Pro Arg Gly Ala
    1910   1915   1920

tta ctc ttt atg gcc tgg ctg ggc gtg aga gtg ctc cct atc aca   5814
Leu Leu Phe Met Ala Trp Leu Gly Val Arg Val Leu Pro Ile Thr
    1925   1930   1935

agg cag ttg gac ggg ggt aac cag gag caa cga ata ccc ggt agc   5859
Arg Gln Leu Asp Gly Gly Asn Gln Glu Gln Arg Ile Pro Gly Ser
    1940   1945   1950

aca aaa ccg aat gcc gaa aat gtg gtc acc gtt tac ggt gca tgg   5904

42

Thr Lys Pro Asn Ala Glu Asn Val Val Thr Val Tyr Gly Ala Trp
  1955   1960   1965

ccc acg ccg cgt agt cca ctg ctg cac ttt gct cca aat gct acg   5949
Pro Thr Pro Arg Ser Pro Leu Leu His Phe Ala Pro Asn Ala Thr
  1970   1975   1980

gag gag ctg gac cag tta cta agc ggc ttt gag gag ttt gag gag   5994
Glu Glu Leu Asp Gln Leu Leu Ser Gly Phe Glu Glu Phe Glu Glu
  1985   1990   1995

agt ttg gga tac aag ttc cgg gat cgg tcg tac ctg ttg caa gcc   6039
Ser Leu Gly Tyr Lys Phe Arg Asp Arg Ser Tyr Leu Leu Gln Ala
  2000   2005   2010

atg aca cat gcc agt tac acg ccc aat cga ttg acg gat tgc tat   6084
Met Thr His Ala Ser Tyr Thr Pro Asn Arg Leu Thr Asp Cys Tyr
  2015   2020   2025

cag cgt ctg gag ttc ctg ggc gat gct gtt cta gat tac ctc att   6129
Gln Arg Leu Glu Phe Leu Gly Asp Ala Val Leu Asp Tyr Leu Ile
  2030   2035   2040

acg cgg cat tta tac gaa gat ccc cgc cag cat tct cca ggc gca   6174
Thr Arg His Leu Tyr Glu Asp Pro Arg Gln His Ser Pro Gly Ala
  2045   2050   2055

tta acg gat ttg cgg tca gca ctg gtg aat aat aca ata ttc gcc   6219
Leu Thr Asp Leu Arg Ser Ala Leu Val Asn Asn Thr Ile Phe Ala
  2060   2065   2070

tcc ctg gct gtt cgc cat ggc ttc cac aag ttc ttc cgg cac ctc   6264
Ser Leu Ala Val Arg His Gly Phe His Lys Phe Phe Arg His Leu
  2075   2080   2085

tcg ccg ggc ctt aac gat gtg att gac cgt ttt gtg cgg atc cag   6309
Ser Pro Gly Leu Asn Asp Val Ile Asp Arg Phe Val Arg Ile Gln

```
          2090              2095              2100


cag gag  aat gga cac tgc atc  agt gag gag tac tac  tta ttg tcc      6354
Gln Glu  Asn Gly His Cys Ile  Ser Glu Glu Tyr Tyr  Leu Leu Ser
     2105              2110              2115


gag gag  gag tgc gat gac gcc  gag gac gtt gag gtg  ccc aag gca      6399
Glu Glu  Glu Cys Asp Asp Ala  Glu Asp Val Glu Val  Pro Lys Ala
     2120              2125              2130


ttg ggc  gac gtt ttc gag tcg  atc gca ggt gcc att  ttt ctc gac      6444
Leu Gly  Asp Val Phe Glu Ser  Ile Ala Gly Ala Ile  Phe Leu Asp
     2135              2140              2145  ·


tca aac  atg tcg ctg gac gtg  gtt tgg cac gta tat  agc aac atg      6489
Ser Asn  Met Ser Leu Asp Val  Val Trp His Val Tyr  Ser Asn Met
     2150              2155              2160


atg agc  ccg gag atc gag cag  ttc agc aac tca gtg  cca aaa tcg      6534
Met Ser  Pro Glu Ile Glu Gln  Phe Ser Asn Ser Val  Pro Lys Ser
     2165              2170              2175


ccc att  cgg gag ctc ctc gag  ctg gag ccg gaa acc  gcc aag ttc      6579
Pro Ile  Arg Glu Leu Leu Glu  Leu Glu Pro Glu Thr  Ala Lys Phe
     2180              2185              2190


ggc aag  ccc gag aag ctg gcg  gat ggg cga cgg gtg  cgc gtt acc      6624
Gly Lys  Pro Glu Lys Leu Ala  Asp Gly Arg Arg Val  Arg Val Thr
     2195              2200              2205


gtg gat  gtc ttc tgc aaa gga  acc ttc cgt ggc atc  gga cgc aac      6669
Val Asp  Val Phe Cys Lys Gly  Thr Phe Arg Gly Ile  Gly Arg Asn
     2210              2215              2220


tat cgc  att gcc aag tgc acg  gcg gcc aaa tgc gca  ttg cgc caa      6714
Tyr Arg  Ile Ala Lys Cys Thr  Ala Ala Lys Cys Ala  Leu Arg Gln
     2225              2230              2235
```

ctc aaa aag cag ggc ttg ata gcc aaa aaa gac taa   6750
Leu Lys Lys Gln Gly Leu Ile Ala Lys Lys Asp
    2240   2245

<210> 4
<211> 2249
<212> PRT
<213> Drosophila melanogaster

<400> 4

Met Ala Phe His Trp Cys Asp Asn Asn Leu His Thr Thr Val Phe Thr
1   5   10   15

Pro Arg Asp Phe Gln Val Glu Leu Leu Ala Thr Ala Tyr Glu Arg Asn
        20   25   30

Thr Ile Ile Cys Leu Gly His Arg Ser Ser Lys Glu Phe Ile Ala Leu
        35   40   45

Lys Leu Leu Gln Glu Leu Ser Arg Arg Ala Arg Arg His Gly Arg Val
    50   55   60

Ser Val Tyr Leu Ser Cys Glu Val Gly Thr Ser Thr Glu Pro Cys Ser
65   70   75   80

Ile Tyr Thr Met Leu Thr His Leu Thr Asp Leu Arg Val Trp Gln Glu
            85   90   95

Gln Pro Asp Met Gln Ile Pro Phe Asp His Cys Trp Thr Asp Tyr His
      100   105   110

Val Ser Ile Leu Arg Pro Glu Gly Phe Leu Tyr Leu Leu Glu Thr Arg
      115   120   125

Glu Leu Leu Leu Ser Ser Val Glu Leu Ile Val Leu Glu Asp Cys His
      130   135   140

Asp Ser Ala Val Tyr Gln Arg Ile Arg Pro Leu Phe Glu Asn His Ile
   145   150   155   160

Met Pro Ala Pro Pro Ala Asp Arg Pro Arg Ile Leu Gly Leu Ala Gly
      165   170   175

Pro Leu His Ser Ala Gly Cys Glu Leu Gln Gln Leu Ser Ala Met Leu
      180   185   190

Ala Thr Leu Glu Gln Ser Val Leu Cys Gln Ile Glu Thr Ala Ser Asp
      195   200   205

Ile Val Thr Val Leu Arg Tyr Cys Ser Arg Pro His Glu Tyr Ile Val
      210   215   220

Gln Cys Ala Pro Phe Glu Met Asp Glu Leu Ser Leu Val Leu Ala Asp
   225   230   235   240

Val Leu Asn Thr His Lys Ser Phe Leu Leu Asp His Arg Tyr Asp Pro

```
                  245            250            255
```

```
Tyr Glu Ile Tyr Gly Thr Asp Gln Phe Met Asp Glu Leu Lys Asp Ile
          260            265            270
```

```
Pro Asp Pro Lys Val Asp Pro Leu Asn Val Ile Asn Ser Leu Leu Val
          275            280            285
```

```
Val Leu His Glu Met Gly Pro Trp Cys Thr Gln Arg Ala Ala His His
          290            295            300
```

```
Phe Tyr Gln Cys Asn Glu Lys Leu Lys Val Lys Thr Pro His Glu Arg
305            310            315            320
```

```
His Tyr Leu Leu Tyr Cys Leu Val Ser Thr Ala Leu Ile Gln Leu Tyr
               325            330            335
```

```
Ser Leu Cys Glu His Ala Phe His Arg His Leu Gly Ser Gly Ser Asp
          340            345            350
```

```
Ser Arg Gln Thr Ile Glu Arg Tyr Ser Ser Pro Lys Val Arg Arg Leu
          355            360            365
```

```
Leu Gln Thr Leu Arg Cys Phe Lys Pro Glu Glu Val His Thr Gln Ala
          370            375            380
```

```
Asp Gly Leu Arg Arg Met Arg His Gln Val Asp Gln Ala Asp Phe Asn
385            390            395            400
```

Arg Leu Ser His Thr Leu Glu Ser Lys Cys Arg Met Val Asp Gln Met
          405   410   415

Asp Gln Pro Pro Thr Glu Thr Arg Ala Leu Val Ala Thr Leu Glu Gln
          420   425   430

Ile Leu His Thr Thr Glu Asp Arg Gln Thr Asn Arg Ser Ala Ala Arg
          435   440   445

Val Thr Pro Thr Pro Thr Pro Ala His Ala Lys Pro Lys Pro Ser Ser
        450   455   460

Gly Ala Asn Thr Ala Gln Pro Arg Thr Arg Arg Arg Val Tyr Thr Arg
465   470   475   480

Arg His His Arg Asp His Asn Asp Gly Ser Asp Thr Leu Cys Ala Leu
          485   490   495

Ile Tyr Cys Asn Gln Asn His Thr Ala Arg Val Leu Phe Glu Leu Leu
          500   505   510

Ala Glu Ile Ser Arg Arg Asp Pro Asp Leu Lys Phe Leu Arg Cys Gln
          515   520   525

Tyr Thr Thr Asp Arg Val Ala Asp Pro Thr Thr Glu Pro Lys Glu Ala
        530   535   540

Glu Leu Glu His Arg Arg Gln Glu Glu Val Leu Lys Arg Phe Arg Met
545   550   555   560

His Asp Cys Asn Val Leu Ile Gly Thr Ser Val Leu Glu Glu Gly Ile
        565   570   575

Asp Val Pro Lys Cys Asn Leu Val Val Arg Trp Asp Pro Pro Thr Thr
        580   585   590

Tyr Arg Ser Tyr Val Gln Cys Lys Gly Arg Ala Arg Ala Ala Pro Ala
        595   600   605

Tyr His Val Ile Leu Val Ala Pro Ser Tyr Lys Ser Pro Thr Val Gly
        610   615   620

Ser Val Gln Leu Thr Asp Arg Ser His Arg Tyr Ile Cys Ala Thr Gly
625   630   635   640

Asp Thr Thr Glu Ala Asp Ser Asp Ser Asp Asp Ser Ala Met Pro Asn
        645   650   655

Ser Ser Gly Ser Asp Pro Tyr Thr Phe Gly Thr Ala Arg Gly Thr Val
        660   665   670

Lys Ile Leu Asn Pro Glu Val Phe Ser Lys Gln Pro Pro Thr Ala Cys
        675   680   685

Asp Ile Lys Leu Gln Glu Ile Gln Asp Glu Leu Pro Ala Ala Ala Gln
    690   695   700

Leu Asp Thr Ser Asn Ser Ser Asp Glu Ala Val Ser Met Ser Asn Thr
705   710   715   720

Ser Pro Ser Glu Ser Ser Thr Glu Gln Lys Ser Arg Arg Phe Gln Cys
             725   730   735

Glu Leu Ser Ser Leu Thr Glu Pro Glu Asp Thr Ser Asp Thr Thr Ala
             740   745   750

Glu Ile Asp Thr Ala His Ser Leu Ala Ser Thr Thr Lys Asp Leu Val
             755   760   765

His Gln Met Ala Gln Tyr Arg Glu Ile Glu Gln Met Leu Leu Ser Lys
    770   775   780

Cys Ala Asn Thr Glu Pro Pro Glu Gln Glu Gln Ser Glu Ala Glu Arg
785   790   795   800

Phe Ser Ala Cys Leu Ala Ala Tyr Arg Pro Lys Pro His Leu Leu Thr
             805   810   815

Gly Ala Ser Val Asp Leu Gly Ser Ala Ile Ala Leu Val Asn Lys Tyr
             820   825   830

Cys Ala Arg Leu Pro Ser Asp Thr Phe Thr Lys Leu Thr Ala Leu Trp

        835   840   845

Arg Cys Thr Arg Asn Glu Arg Ala Gly Val Thr Leu Phe Gln Tyr Thr
     850   855   860

Leu Arg Leu Pro Ile Asn Ser Pro Leu Lys His Asp Ile Val Gly Leu
865   870   875   880

Pro Met Pro Thr Gln Thr Leu Ala Arg Arg Leu Ala Ala Leu Gln Ala
          885   890   895

Cys Val Glu Leu His Arg Ile Gly Glu Leu Asp Asp Gln Leu Gln Pro
          900   905   910

Ile Gly Lys Glu Gly Phe Arg Ala Leu Glu Pro Asp Trp Glu Cys Phe
          915   920   925

Glu Leu Glu Pro Glu Asp Glu Gln Ile Val Gln Leu Ser Asp Glu Pro
     930   935   940

Arg Pro Gly Thr Thr Lys Arg Arg Gln Tyr Tyr Tyr Lys Arg Ile Ala
945   950   955   960

Ser Glu Phe Cys Asp Cys Arg Pro Val Ala Gly Ala Pro Cys Tyr Leu
          965   970   975

Tyr Phe Ile Gln Leu Thr Leu Gln Cys Pro Ile Pro Glu Glu Gln Asn
          980   985   990

Thr Arg Gly Arg Lys Ile Tyr Pro Pro Glu Asp Ala Gln Gln Gly Phe
 995   1000   1005

Gly Ile Leu Thr Thr Lys Arg Ile Pro Lys Leu Ser Ala Phe Ser
 1010   1015   1020

Ile Phe Thr Arg Ser Gly Glu Val Lys Val Ser Leu Glu Leu Ala
 1025   1030   1035

Lys Glu Arg Val Ile Leu Thr Ser Glu Gln Ile Val Cys Ile Asn
 1040   1045   1050

Gly Phe Leu Asn Tyr Thr Phe Thr Asn Val Leu Arg Leu Gln Lys
 1055   1060   1065

Phe Leu Met Leu Phe Asp Pro Asp Ser Thr Glu Asn Cys Val Phe
 1070   1075   1080

Ile Val Pro Thr Val Lys Ala Pro Ala Gly Gly Lys His Ile Asp
 1085   1090   1095

Trp Gln Phe Leu Glu Leu Ile Gln Ala Asn Gly Asn Thr Met Pro
 1100   1105   1110

Arg Ala Val Pro Asp Glu Glu Arg Gln Ala Gln Pro Phe Asp Pro
 1115   1120   1125

Gln Arg Phe Gln Asp Ala Val Val Met Pro Trp Tyr Arg Asn Gln
   1130   1135   1140

Asp Gln Pro Gln Tyr Phe Tyr Val Ala Glu Ile Cys Pro His Leu
   1145   1150   1155

Ser Pro Leu Ser Cys Phe Pro Gly Asp Asn Tyr Arg Thr Phe Lys
   1160   1165   1170

His Tyr Tyr Leu Val Lys Tyr Gly Leu Thr Ile Gln Asn Thr Ser
   1175   1180   1185

Gln Pro Leu Leu Asp Val Asp His Thr Ser Ala Arg Leu Asn Phe
   1190   1195   1200

Leu Thr Pro Arg Tyr Val Asn Arg Lys Gly Val Ala Leu Pro Thr
   1205   1210   1215

Ser Ser Glu Glu Thr Lys Arg Ala Lys Arg Glu Asn Leu Glu Gln
   1220   1225   1230

Lys Gln Ile Leu Val Pro Glu Leu Cys Thr Val His Pro Phe Pro
   1235   1240   1245

Ala Ser Leu Trp Arg Thr Ala Val Cys Leu Pro Cys Ile Leu Tyr
   1250   1255   1260

```
Arg Ile  Asn Gly Leu Leu Leu  Ala Asp Asp Ile Arg  Lys Gln Val
    1265          1270         1275


Ser Ala  Asp Leu Gly Leu Gly  Arg Gln Gln Ile Glu  Asp Glu Asp
    1280          1285         1290


Phe Glu  Trp Pro Met Leu Asp  Phe Gly Trp Ser Leu  Ser Glu Val
    1295          1300         1305


Leu Lys  Lys Ser Arg Glu Ser  Lys Gln Lys Glu Ser  Leu Lys Asp
    1310          1315         1320


Asp Thr  Ile Asn Gly Lys Asp  Leu Ala Asp Val Glu  Lys Lys Pro
    1325          1330         1335


Thr Ser  Glu Glu Thr Gln Leu  Asp Lys Asp Ser Lys  Asp Asp Lys
    1340          1345         1350


Val Glu  Lys Ser Ala Ile Glu  Leu Ile Ile Glu Gly  Glu Glu Lys
    1355          1360         1365


Leu Gln  Glu Ala Asp Asp Phe  Ile Glu Ile Gly Thr  Trp Ser Asn
    1370          1375         1380


Asp Met  Ala Asp Asp Ile Ala  Ser Phe Asn Gln Glu  Asp Asp Asp
    1385          1390         1395


Glu Asp  Asp Ala Phe His Leu  Pro Val Leu Pro Ala  Asn Val Lys
```

1400   1405   1410

Phe Cys Asp Gln Gln Thr Arg Tyr Gly Ser Pro Thr Phe Trp Asp
    1415   1420   1425

Val Ser Asn Gly Glu Ser Gly Phe Lys Gly Pro Lys Ser Ser Gln
    1430   1435   1440

Asn Lys Gln Gly Gly Lys Gly Lys Ala Lys Gly Pro Ala Lys Pro
    1445   1450   1455

Thr Phe Asn Tyr Tyr Asp Ser Asp Asn Ser Leu Gly Ser Ser Tyr
    1460   1465   1470

Asp Asp Asp Asp Asn Ala Gly Pro Leu Asn Tyr Met His His Asn
    1475   1480   1485

Tyr Ser Ser Asp Asp Asp Asp Val Ala Asp Asp Ile Asp Ala Gly
    1490   1495   1500

Arg Ile Ala Phe Thr Ser Lys Asn Glu Ala Glu Thr Ile Glu Thr
    1505   1510   1515

Ala Gln Glu Val Glu Lys Arg Gln Lys Gln Leu Ser Ile Ile Gln
    1520   1525   1530

Ala Thr Asn Ala Asn Glu Arg Gln Tyr Gln Gln Thr Lys Asn Leu
    1535   1540   1545

Leu Ile Gly Phe Asn Phe Lys His Glu Asp Gln Lys Glu Pro Ala
    1550   1555   1560

Thr Ile Arg Tyr Glu Glu Ser Ile Ala Lys Leu Lys Thr Glu Ile
    1565   1570   1575

Glu Ser Gly Gly Met Leu Val Pro His Asp Gln Gln Leu Val Leu
    1580   1585   1590

Lys Arg Ser Asp Ala Ala Glu Ala Gln Val Ala Lys Val Ser Met
    1595   1600   1605

Met Glu Leu Leu Lys Gln Leu Leu Pro Tyr Val Asn Glu Asp Val
    1610   1615   1620

Leu Ala Lys Lys Leu Gly Asp Arg Arg Glu Leu Leu Leu Ser Asp
    1625   1630   1635

Leu Val Glu Leu Asn Ala Asp Trp Val Ala Arg His Glu Gln Glu
    1640   1645   1650

Thr Tyr Asn Val Met Gly Cys Gly Asp Ser Phe Asp Asn Tyr Asn
    1655   1660   1665

Asp His His Arg Leu Asn Leu Asp Glu Lys Gln Leu Lys Leu Gln
    1670   1675   1680

Tyr Glu Arg Ile Glu Ile Glu Pro Pro Thr Ser Thr Lys Ala Ile
1685   1690   1695

Thr Ser Ala Ile Leu Pro Ala Gly Phe Ser Phe Asp Arg Gln Pro
1700   1705   1710

Asp Leu Val Gly His Pro Gly Pro Ser Pro Ser Ile Ile Leu Gln
1715   1720   1725

Ala Leu Thr Met Ser Asn Ala Asn Asp Gly Ile Asn Leu Glu Arg
1730   1735   1740

Leu Glu Thr Ile Gly Asp Ser Phe Leu Lys Tyr Ala Ile Thr Thr
1745   1750   1755

Tyr Leu Tyr Ile Thr Tyr Glu Asn Val His Glu Gly Lys Leu Ser
1760   1765   1770

His Leu Arg Ser Lys Gln Val Ala Asn Leu Asn Leu Tyr Arg Leu
1775   1780   1785

Gly Arg Arg Lys Arg Leu Gly Glu Tyr Met Ile Ala Thr Lys Phe
1790   1795   1800

Glu Pro His Asp Asn Trp Leu Pro Pro Cys Tyr Tyr Val Pro Lys
1805   1810   1815

Glu Leu Glu Lys Ala Leu Ile Glu Ala Lys Ile Pro Thr His His
      1820   1825   1830

Trp Lys Leu Ala Asp Leu Leu Asp Ile Lys Asn Leu Ser Ser Val
      1835   1840   1845

Gln Ile Cys Glu Met Val Arg Glu Lys Ala Asp Ala Leu Gly Leu
      1850   1855   1860

Glu Gln Asn Gly Gly Ala Gln Asn Gly Gln Leu Asp Asp Ser Asn
      1865   1870   1875

Asp Ser Cys Asn Asp Phe Ser Cys Phe Ile Pro Tyr Asn Leu Val
      1880   1885   1890

Ser Gln His Ser Ile Pro Asp Lys Ser Ile Ala Asp Cys Val Glu
      1895   1900   1905

Ala Leu Ile Gly Ala Tyr Leu Ile Glu Cys Gly Pro Arg Gly Ala
      1910   1915   1920

Leu Leu Phe Met Ala Trp Leu Gly Val Arg Val Leu Pro Ile Thr
      1925   1930   1935

Arg Gln Leu Asp Gly Gly Asn Gln Glu Gln Arg Ile Pro Gly Ser
      1940   1945   1950

Thr Lys Pro Asn Ala Glu Asn Val Val Thr Val Tyr Gly Ala Trp

                1955   1960   1965

Pro Thr Pro Arg Ser Pro Leu Leu His Phe Ala Pro Asn Ala Thr
    1970   1975   1980

Glu Glu Leu Asp Gln Leu Leu Ser Gly Phe Glu Glu Phe Glu Glu
    1985   1990   1995

Ser Leu Gly Tyr Lys Phe Arg Asp Arg Ser Tyr Leu Leu Gln Ala
    2000   2005   2010

Met Thr His Ala Ser Tyr Thr Pro Asn Arg Leu Thr Asp Cys Tyr
    2015   2020   2025

Gln Arg Leu Glu Phe Leu Gly Asp Ala Val Leu Asp Tyr Leu Ile
    2030   2035   2040

Thr Arg His Leu Tyr Glu Asp Pro Arg Gln His Ser Pro Gly Ala
    2045   2050   2055

Leu Thr Asp Leu Arg Ser Ala Leu Val Asn Asn Thr Ile Phe Ala
    2060   2065   2070

Ser Leu Ala Val Arg His Gly Phe His Lys Phe Phe Arg His Leu
    2075   2080   2085

Ser Pro Gly Leu Asn Asp Val Ile Asp Arg Phe Val Arg Ile Gln
    2090   2095   2100

Gln Glu Asn Gly His Cys Ile Ser Glu Glu Tyr Tyr Leu Leu Ser
  2105   2110   2115

Glu Glu Glu Cys Asp Asp Ala Glu Asp Val Glu Val Pro Lys Ala
  2120   2125   2130

Leu Gly Asp Val Phe Glu Ser Ile Ala Gly Ala Ile Phe Leu Asp
  2135   2140   2145

Ser Asn Met Ser Leu Asp Val Val Trp His Val Tyr Ser Asn Met
  2150   2155   2160

Met Ser Pro Glu Ile Glu Gln Phe Ser Asn Ser Val Pro Lys Ser
  2165   2170   2175

Pro Ile Arg Glu Leu Leu Glu Leu Glu Pro Glu Thr Ala Lys Phe
  2180   2185   2190

Gly Lys Pro Glu Lys Leu Ala Asp Gly Arg Arg Val Arg Val Thr
  2195   2200   2205

Val Asp Val Phe Cys Lys Gly Thr Phe Arg Gly Ile Gly Arg Asn
  2210   2215   2220

Tyr Arg Ile Ala Lys Cys Thr Ala Ala Lys Cys Ala Leu Arg Gln
  2225   2230   2235

Leu Lys Lys Gln Gly Leu Ile Ala Lys Lys Asp
   2240   2245

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 June 2002 (06.06.2002)

PCT

(10) International Publication Number
WO 02/44321 A2

(51) International Patent Classification[7]:   **C12N**

(21) International Application Number: PCT/EP01/13968

(22) International Filing Date:
29 November 2001 (29.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
00126325.0	1 December 2000 (01.12.2000)	EP
60/279,661	30 March 2001 (30.03.2001)	US
PCT/US01/10188	30 March 2001 (30.03.2001)	US

(71) Applicant *(for all designated States except US)*: **MAX-PLANCK-GESELLSCHAFT ZUR FÖRDERUNG DER WISSENSCHAFTEN E.V.** [DE/DE]; Hofgartenstrasse 2, 80539 München (DE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **TUSCHL, Thomas** [DE/DE]; Keplerstrasse 9, 37085 Göttingen (DE). **EL-BASHIR, Sayda** [DE/DE]; Sultebecksbreite 2, 37075 Göttingen (DE). **LENDECKEL, Winfried** [DE/DE]; Blinde Gasse 52, 37318 Hohengandern (DE).

(74) **Agent: WEICKMANN & WEICKMANN;** Postfach 860 820, 81635 München (DE).

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** RNA INTERFERENCE MEDIATING SMALL RNA MOLECULES

(57) **Abstract:** Double-stranded RNA (dsRNA) induces sequence-specific post-transcriptional gene silencing in many organisms by a process known as RNA interference (RNAi). Using a Drosophila in vitro system, we demonstrate that 19-23 nt short RNA fragments are the sequence-specific mediators of RNAi. The short interfering RNAs (siRNAs) are generated by an RNase III-like processing reaction from long dsRNA. Chemically synthesized siRNA duplexes with overhanging 3' ends mediate efficient target RNA cleavage in the lysate, and the cleavage site is located near the center of the region spanned by the guiding siRNA. Furthermore, we provide evidence that the direction of dsRNA processing determines whether sense or antisense target RNA can be cleaved by the produced siRNP complex.

RNA Interference Mediating Small RNA molecules

Description

The present invention relates to sequence and structural features of double-stranded (ds)RNA molecules required to mediate target-specific nucleic acid modifications such as RNA-interference and/or DNA methylation.

The term "RNA interference" (RNAi) was coined after the discovery that injection of dsRNA into the nematode C. elegans leads to specific silencing of genes highly homologous in sequence to the delivered dsRNA (Fire et al., 1998). RNAi was subsequently also observed in insects, frogs (Oelgeschlager et al., 2000), and other animals including mice (Svoboda et al., 2000; Wianny and Zernicka-Goetz, 2000) and is likely to also exist in human. RNAi is closely linked to the post-transcriptional gene-silencing (PTGS) mechanism of co-suppression in plants and quelling in fungi (Catalanotto et al., 2000; Cogoni and Macino, 1999; Dalmay et al., 2000; Ketting and Plasterk, 2000; Mourrain et al., 2000; Smardon et al., 2000) and some components of the RNAi machinery are also necessary for post-transcriptional silencing by co-suppression (Catalanotto et al., 2000; Dernburg et al., 2000; Ketting and Plasterk, 2000). The topic has also been reviewed recently (Bass, 2000; Bosher and Labouesse, 2000; Fire, 1999; Plasterk and Ketting, 2000; Sharp, 1999; Sijen and Kooter, 2000), see also the entire issue of Plant Molecular Biology, vol. 43, issue 2/3, (2000).

In plants, in addition to PTGS, introduced transgenes can also lead to transcriptional gene silencing via RNA-directed DNA methylation of cytosines (see references in Wassenegger, 2000). Genomic targets as short as 30 bp are methylated in plants in an RNA-directed manner (Pelissier, 2000). DNA methylation is also present in mammals.

- 2 -

The natural function of RNAi and co-suppression appears to be protection of the genome against invasion by mobile genetic elements such as retrotransposons and viruses which produce aberrant RNA or dsRNA in the host cell when they become active (Jensen et al, 1999; Ketting et al., 1999;
5   Ratcliff et al., 1999; Tabara et al., 1999). Specific mRNA degradation prevents transposon and virus replication although some viruses are able to overcome or prevent this process by expressing proteins that suppress PTGS (Lucy et al., 2000; Voinnet et al., 2000).

10   DsRNA triggers the specific degradation of homologous RNAs only within the region of identity with the dsRNA (Zamore et al., 2000). The dsRNA is processed to 21-23 nt RNA fragments and the target RNA cleavage sites are regularly spaced 21-23 nt apart. It has therefore been suggested that the 21-23 nt fragments are the guide RNAs for target recognition (Zamore
15   et al., 2000). These short RNAs were also detected in extracts prepared from D. melanogaster Schneider 2 cells which were transfected with dsRNA prior to cell lysis (Hammond et al., 2000), however, the fractions that displayed sequence-specific nuclease activity also contained a large fraction of residual dsRNA. The role of the 21-23 nt fragments in guiding
20   mRNA cleavage is further supported by the observation that 21-23 nt fragments isolated from processed dsRNA are able, to some extent, to mediate specific mRNA degradation (Zamore et al., 2000). RNA molecules of similar size also accumulate in plant tissue that exhibits PTGS (Hamilton and Baulcombe, 1999).

25

Here, we use the established Drosophila in vitro system (Tuschl et al., 1999; Zamore et al., 2000) to further explore the mechanism of RNAi. We demonstrate that short 21 and 22 nt RNAs, when base-paired with 3' overhanging ends, act as the guide RNAs for sequence-specific mRNA
30   degradation. Short 30 bp dsRNAs are unable to mediate RNAi in this system because they are no longer processed to 21 and 22 nt RNAs. Furthermore, we defined the target RNA cleavage sites relative to the 21 and

22 nt short interfering RNAs (siRNAs) and provide evidence that the direction of dsRNA processing determines whether a sense or an antisense target RNA can be cleaved by the produced siRNP endonuclease complex. Further, the siRNAs may also be important tools for transcriptional modulating, e.g. silencing of mammalian genes by guiding DNA methylation.

Further experiments in human in vivo cell culture systems (HeLa cells) show that double-stranded RNA molecules having a length of preferably from 19-25 nucleotides have RNAi activity. Thus, in contrast to the results from Drosophila also 24 and 25 nt long double-stranded RNA molecules are efficient for RNAi.

The object underlying the present invention is to provide novel agents capable of mediating target-specific RNA interference or other target-specific nucleic acid modifications such as DNA methylation, said agents having an improved efficacy and safety compared to prior art agents.

The solution of this problem is provided by an isolated double-stranded RNA molecule, wherein each RNA strand has a length from 19-25, particularly from 19-23 nucleotides, wherein said RNA molecule is capable of mediating target-specific nucleic acid modifications, particularly RNA interference and/or DNA methylation. Preferably at least one strand has a 3'-overhang from 1-5 nucleotides, more preferably from 1-3 nucleotides and most preferably 2 nucleotides. The other strand may be blunt-ended or has up to 6 nucleotides 3' overhang. Also, if both strands of the dsRNA are exactly 21 or 22 nt, it is possible to observe some RNA interference when both ends are blunt (0 nt overhang). The RNA molecule is preferably a synthetic RNA molecule which is substantially free from contaminants occurring in cell extracts, e.g. from Drosophila embryos. Further, the RNA molecule is preferably substantially free from any non-target-specific contaminants, particularly non-target-specific RNA molecules e.g. from contaminants occuring in cell extracts.

- 4 -

Further, the invention relates to the use of isolated double-stranded RNA molecules, wherein each RNA strand has a length from 19-25 nucleotides, for mediating, target-specific nucleic acid modifications, particularly RNAi, in mammalian cells, particularly in human cells.

5

Surprisingly, it was found that synthetic short double-stranded RNA molecules particularly with overhanging 3'-ends are sequence-specific mediators of RNAi and mediate efficient target-RNA cleavage, wherein the cleavage site is located near the center of the region spanned by the guiding short
10   RNA.

Preferably, each strand of the RNA molecule has a length from 20-22 nucleotides (or 20-25 nucleotides in mammalian cells), wherein the length of each strand may be the same or different. Preferably, the length of the
15   3'-overhang reaches from 1-3 nucleotides, wherein the length of the overhang may be the same or different for each strand. The RNA-strands preferably have 3'-hydroxyl groups. The 5'-terminus preferably comprises a phosphate, diphosphate, triphosphate or hydroxyl group. The most effective dsRNAs are composed of two 21 nt strands which are paired
20   such that 1-3, particularly 2 nt 3' overhangs are present on both ends of the dsRNA.

The target RNA cleavage reaction guided by siRNAs is highly sequence-specific. However, not all positions of a siRNA contribute equally to target
25   recognition. Mismatches in the center of the siRNA duplex are most critical and essentially abolish target RNA cleavage. In contrast, the 3' nucleotide of the siRNA strand (e.g. position 21) that is complementary to the single-stranded target RNA, does not contribute to specificity of the target recognition. Further, the sequence of the unpaired 2-nt 3' overhang of the
30   siRNA strand with the same polarity as the target RNA is not critical for target RNA cleavage as only the antisense siRNA strand guides target recognition. Thus, from the single-stranded overhanging nucleotides only the

- 5 -

penultimate position of the antisense siRNA (e.g. position 20) needs to match the targeted sense mRNA.

Surprisingly, the double-stranded RNA molecules of the present invention exhibit a high in vivo stability in serum or in growth medium for cell cultures. In order to further enhance the stability, the 3'-overhangs may be stablized against degradation, e.g. they may be selected such that they consist of purine nucleotides, particularly adenosine or guanosine nucleotides. Alternatively, substitution of pyrimidine nucleotides by modified analogues, e.g. substitution of uridine 2 nt 3' overhangs by 2'-deoxythymidine is tolerated and does not affect the efficiency of RNA interference. The absence of a 2' hydroxyl significantly enhances the nuclease resistance of the overhang in tissue culture medium.

In an especially preferred embodiment of the present invention the RNA molecule may contain at least one modified nucleotide analogue. The nucleotide analogues may be located at positions where the target-specific activity, e.g. the RNAi mediating activity is not substantially effected, e.g. in a region at the 5'-end and/or the 3'-end of the double-stranded RNA molecule. Particularly, the overhangs may be stabilized by incorporating modified nucleotide analogues.

Preferred nucleotide analogues are selected from sugar- or backbone-modified ribonucleotides. It should be noted, however, that also nucleobase-modified ribonucleotides, i.e. ribonucleotides, containing a non-naturally occurring nucleobase instead of a naturally occurring nucleobase such as uridines or cytidines modified at the 5-position, e.g. 5-(2-amino)propyl uridine, 5-bromo uridine; adenosines and guanosines modified at the 8-position, e.g. 8-bromo guanosine; deaza nucleotides, e.g. 7-deaza-adenosine; O- and N-alkylated nucleotides, e.g. N6-methyl adenosine are suitable. In preferred sugar-modified ribonucleotides the 2´ OH-group is repla-

- 6 -

ced by a group selected from H, OR, R, halo, SH, SR, NH$_2$, NHR, NR$_2$ or CN, wherein R is C$_1$-C$_6$ alkyl, alkenyl or alkynyl and halo is F, Cl, Br or I. In preferred backbone-modified ribonucleotides the phosphoester group connecting to adjacent ribonucleotides is replaced by a modified group, e.g. of phosphothioate group. It should be noted that the above modifications may be combined.

The sequence of the double-stranded RNA molecule of the present invention has to have a sufficient identity to a nucleic acid target molecule in order to mediate target-specific RNAi and/or DNA methylation. Preferably, the sequence has an identity of at least 50%, particularly of at least 70% to the desired target molecule in the double-stranded portion of the RNA molecule. More preferably, the identity is at least 85% and most preferably 100% in the double-stranded portion of the RNA molecule. The identity of a double-stranded RNA molecule to a predetermined nucleic acid target molecule, e.g. an mRNA target molecule may be determined as follows:

$$I = \frac{n}{L} \times 100$$

wherein I is the identity in percent, n is the number of identical nucleotides in the double-stranded portion of the ds RNA and the target and L is the length of the sequence overlap of the double-stranded portion of the dsRNA and the target.

Alternatively, the identity of the double-stranded RNA molecule to the target sequence may also be defined including the 3' overhang, particularly an overhang having a length from 1-3 nucleotides. In this case the sequence identity is preferably at least 50%, more preferably at least 70% and most preferably at least 85% to the target sequence. For example, the nucleotides from the 3' overhang and up to 2 nucleotides from the 5' and/or 3' terminus of the double strand may be modified without significant loss of activity.

The double-stranded RNA molecule of the invention may be prepared by a method comprising the steps:

(a)   synthesizing two RNA strands each having a length from 19-25, e.g. from 19-23 nucleotides, wherein said RNA strands are capable of forming a double-stranded RNA molecule, wherein preferably at least one strand has a 3'-overhang from 1-5 nucleotides,

(b)   combining the synthesized RNA strands under conditions, wherein a double-stranded RNA molecule is formed, which is capable of mediating target-specific nucleic acid modifications, particularly RNA interference and/or DNA methylation.

Methods of synthesizing RNA molecules are known in the art. In this context, it is particularly referred to chemical synthesis methods as described in Verma and Eckstein (1998).

The single-stranded RNAs can also be prepared by enzymatic transcription from synthetic DNA templates or from DNA plasmids isolated from recombinant bacteria. Typically, phage RNA polymerases are used such as T7, T3 or SP6 RNA polymerase (Milligan and Uhlenbeck (1989)).

A further aspect of the present invention relates to a method of mediating target-specific nucleic acid modifications, particularly RNA interference and/or DNA methylation in a cell or an organism comprising the steps:

(a)   contacting the cell or organism with the double-stranded RNA molecule of the invention under conditions wherein target-specific nucleic acid modifications may occur and

(b)   mediating a target-specific nucleic acid modificiation effected by the double-stranded RNA towards a target nucleic acid having a

- 8 -

sequence portion substantially corresponding to the double-stranded RNA.

Preferably the contacting step (a) comprises introducing the double-stranded RNA molecule into a target cell, e.g. an isolated target cell, e.g. in cell culture, a unicellular microorganism or a target cell or a plurality of target cells within a multicellular organism. More preferably, the introducing step comprises a carrier-mediated delivery, e.g. by liposomal carriers or by injection.

The method of the invention may be used for determining the function of a gene in a cell or an organism or even for modulating the function of a gene in a cell or an organism, being capable of mediating RNA interference. The cell is preferably a eukaryotic cell or a cell line, e.g. a plant cell or an animal cell, such as a mammalian cell, e.g. an embryonic cell, a pluripotent stem cell, a tumor cell, e.g. a teratocarcinoma cell or a virus-infected cell. The organism is preferably a eukaryotic organism, e.g. a plant or an animal, such as a mammal, particularly a human.

The target gene to which the RNA molecule of the invention is directed may be associated with a pathological condition. For example, the gene may be a pathogen-associated gene, e.g. a viral gene, a tumor-associated gene or an autoimmune disease-associated gene. The target gene may also be a heterologous gene expressed in a recombinant cell or a genetically altered organism. By determinating or modulating, particularly, inhibiting the function of such a gene valuable information and therapeutic benefits in the agricultural field or in the medicine or veterinary medicine field may be obtained.

The dsRNA is usually administered as a pharmaceutical composition. The administration may be carried out by known methods, wherein a nucleic acid is introduced into a desired target cell in vitro or in vivo. Commonly

used gene transfer techniques include calcium phosphate, DEAE-dextran, electroporation and microinjection and viral methods (Graham, F.L. and van der Eb, A.J. (1973) Virol. 52, 456; McCutchan, J.H. and Pagano, J.S. (1968), J. Natl. Cancer Inst. 41, 351; Chu, G. et al (1987), Nucl. Acids Res. 15, 1311; Fraley, R. et al. (1980), J. Biol. Chem. 255, 10431; Capecchi, M.R. (1980), Cell 22, 479). A recent addition to this arsenal of techniques for the introduction of DNA into cells is the use of cationic liposomes (Felgner, P.L. et al. (1987), Proc. Natl. Acad. Sci USA 84, 7413). Commercially available cationic lipid formulations are e.g. Tfx 50 (Promega) or Lipofectamin2000 (Life Technologies).

Thus, the invention also relates to a pharmaceutical composition containing as an active agent at least one double-stranded RNA molecule as described above and a pharmaceutical carrier. The composition may be used for diagnostic and for therapeutic applications in human medicine or in veterinary medicine.

For diagnostic or therapeutic applications, the composition may be in form of a solution, e.g. an injectable solution, a cream, ointment, tablet, suspension or the like. The composition may be administered in any suitable way, e.g. by injection, by oral, topical, nasal, rectal application etc. The carrier may be any suitable pharmaceutical carrier. Preferably, a carrier is used, which is capable of increasing the efficacy of the RNA molecules to enter the target-cells. Suitable examples of such carriers are liposomes, particularly cationic liposomes. A further preferred administration method is injection.

A further preferred application of the RNAi method is a functional analysis of eukaryotic cells, or eukaryotic non-human organisms, preferably mammalian cells or organisms and most preferably human cells, e.g. cell lines such as HeLa or 293 or rodents, e.g. rats and mice. By transfection with suitable double-stranded RNA molecules which are homologous to a prede-

- 10 -

termined target gene or DNA molecules encoding a suitable double-stranded RNA molecule a specific knockout phenotype can be obtained in a target cell, e.g. in cell culture or in a target organism. Surprisingly it was found that the presence of short double-stranded RNA molecules does not result in an interferon response from the host cell or host organism.

Thus, a further subject matter of the invention is a eukaryotic cell or a eukaryotic non-human organism exhibiting a target gene-specific knockout phenotype comprising an at least partially deficient expression of at least one endogeneous target gene wherein said cell or organism is transfected with at least one double-stranded RNA molecule capable of inhibiting the expression of at least one endogeneous target gene or with a DNA encoding at least one double stranded RNA molecule capable of inhibiting the expression of at least one endogeneous target gene. It should be noted that the present invention allows a target-specific knockout of several different endogeneous genes due to the specificity of RNAi.

Gene-specific knockout phenotypes of cells or non-human organisms, particularly of human cells or non-human mammals may be used in analytic procedures, e.g. in the functional and/or phenotypical analysis of complex physiological processes such as analysis of gene expression profiles and/or proteomes. For example, one may prepare the knock-out phenotypes of human genes in cultured cells which are assumed to be regulators of alternative splicing processes. Among these genes are particularly the members of the SR splicing factor family, e.g. ASF/SF2, SC35, SRp20, SRp40 or SRp55. Further, the effect of SR proteins on the mRNA profiles of predetermined alternatively spliced genes such as CD44 may be analysed. Preferably the analysis is carried out by high-throughput methods using oligonucleotide based chips.

Using RNAi based knockout technologies, the expression of an endogeneous target gene may be inhibited in a target cell or a target organism.

- 11 -

The endogeneous gene may be complemented by an exogeneous target nucleic acid coding for the target protein or a variant or mutated form of the target protein, e.g. a gene or a cDNA, which may optionally be fused to a further nucleic acid sequence encoding a detectable peptide or poly-peptide, e.g. an affinity tag, particularly a multiple affinity tag. Variants or mutated forms of the target gene differ from the endogeneous target gene in that they encode a gene product which differs from the endogeneous gene product on the amino acid level by substitutions, insertions and/or deletions of single or multiple amino acids. The variants or mutated forms may have the same biological activity as the endogeneous target gene. On the other hand, the variant or mutated target gene may also have a biologi-cal activity, which differs from the biological activity of the endogeneous target gene, e.g. a partially deleted activity, a completely deleted activity, an enhanced activity etc.

The complementation may be accomplished by coexpressing the polypep-tide encoded by the exogeneous nucleic acid, e.g. a fusion protein com-prising the target protein and the affinity tag and the double stranded RNA molecule for knocking out the endogeneous gene in the target cell. This coexpression may be accomplished by using a suitable expression vector expressing both the polypeptide encoded by the exogeneous nucleic acid, e.g. the tag-modified target protein and the double stranded RNA molecule or alternatively by using a combination of expression vectors. Proteins and protein complexes which are synthesized de novo in the target cell will contain the exogeneous gene product, e.g. the modified fusion protein. In order to avoid suppression of the exogeneous gene product expression by the RNAi duplex molecule, the nucleotide sequence encoding the exoge-neous nucleic acid may be altered on the DNA level (with or without cau-sing mutations on the amino acid level) in the part of the sequence which is homologous to the double stranded RNA molecule. Alternatively, the endogeneous target gene may be complemented by corresponding nucleo-tide sequences from other species, e.g. from mouse.

- 12 -

Preferred applications for the cell or organism of the invention is the analy-
sis of gene expression profiles and/or proteomes. In an especially preferred
embodiment an analysis of a variant or mutant form of one or several
target proteins is carried out, wherein said variant or mutant forms are
reintroduced into the cell or organism by an exogeneous target nucleic acid
as described above. The combination of knockout of an endogeneous gene
and rescue by using mutated, e.g. partially deleted exogeneous target has
advantages compared to the use of a knockout cell. Further, this method
is particularly suitable for identifying functional domains of the target
protein. In a further preferred embodiment a comparison, e.g. of gene
expression profiles and/or proteomes and/or phenotypic characteristics of
at least two cells or organisms is carried out. These organisms are selected
from:

(i)   a control cell or control organism without target gene inhibition,

(ii)   a cell or organism with target gene inhibition and

(iii)   a cell or organism with target gene inhibition plus target gene com-
         plementation by an exogeneous target nucleic acid.

The method and cell of the invention are also suitable in a procedure for
identifying and/or characterizing pharmacological agents, e.g. identifying
new pharmacological agents from a collection of test substances and/or
characterizing mechanisms of action and/or side effects of known pharma-
cological agents.

Thus, the present invention also relates to a system for identifying and/or
characterizing pharmacological agents acting on at least one target protein
comprising:

(a)   a eukaryotic cell or a eukaryotic non-human organism capable of
         expressing at least one endogeneous target gene coding for said
         target protein,

(b)   at least one double-stranded RNA molecule capable of inhibiting the
         expression of said at least one endogeneous target gene, and

(c)   a test substance or a collection of test substances wherein pharma-
        cological properties of said test substance or said collection are to
        be identified and/or characterized.

Further, the system as described above preferably comprises:

(d)   at least one exogeneous target nucleic acid coding for the target
        protein or a variant or mutated form of the target protein wherein
        said exogeneous target nucleic acid differs from the endogeneous
        target gene on the nucleic acid level such that the expression of the
        exogeneous target nucleic acid is substantially less inhibited by the
        double stranded RNA molecule than the expression of the endoge-
        neous target gene.

Furthermore, the RNA knockout complementation method may be used for
preparative purposes, e.g. for the affinity purification of proteins or protein
complexes from eukaryotic cells, particularly mammalian cells and more
particularly human cells. In this embodiment of the invention, the exoge-
neous target nucleic acid preferably codes for a target protein which is
fused to an affinity tag.

The preparative method may be employed for the purification of high
molecular weight protein complexes which preferably have a mass of $\geq$
150 kD and more preferably of $\geq$ 500 kD and which optionally may con-
tain nucleic acids such as RNA. Specific examples are the heterotrimeric
protein complex consisting of the 20 kD, 60 kD and 90 kD proteins of the
U4/U6 snRNP particle, the splicing factor SF3b from the 17S U2 snRNP
consisting of 5 proteins having molecular weights of 14, 49, 120, 145 and
155 kD and the 25S U4/U6/U5 tri-snRNP particle containing the U4, U5
and U6 snRNA molecules and about 30 proteins, which has a molecular
weight of about 1.7 MD.

- 14 -

This method is suitable for functional proteome analysis in mammalian cells, particularly human cells.

Further, the present invention is explained in more detail in the following figures and examples.

Figure Legends

Figure 1: Double-stranded RNA as short as 38 bp can mediate RNAi.
(A) Graphic representation of dsRNAs used for targeting Pp-luc mRNA. Three series of blunt-ended dsRNAs covering a range of 29 to 504 bp were prepared. The position of the first nucleotide of the sense strand of the dsRNA is indicated relative to the start codon of Pp-luc mRNA (p1). (B) RNA interference assay (Tuschl et al., 1999). Ratios of target Pp-luc to control Rr-luc activity were normalized to a buffer control (black bar). DsRNAs (5 nM) were preincubated in Drosophila lysate for 15 min at 25°C prior to the addition of 7-methyl-guanosine-capped Pp-luc and Rr-luc mRNAs (~50 pM). The incubation was continued for another hour and then analyzed by the dual luciferase assay (Promega). The data are the average from at least four independent experiments ± standard deviation.

Figure 2: A 29 bp dsRNA is no longer processed to 21-23 nt fragments. Time course of 21-23 mer formation from processing of internally ^{32}P-labeled dsRNAs (5 nM) in the Drosophila lysate. The length and source of the dsRNA are indicated. An RNA size marker (M) has been loaded in the left lane and the fragment sizes are indicated. Double bands at time zero are due to incompletely denatured dsRNA.

Figure 3: Short dsRNAs cleave the mRNA target only once.
(A) Denaturing gel electrophoreses of the stable 5' cleavage products produced by 1 h incubation of 10 nM sense or antisense RNA ^{32}P-labeled at the cap with 10 nM dsRNAs of the p133 series in Drosophila lysate.

Length markers were generated by partial nuclease T1 digestion and partial alkaline hydrolysis (OH) of the cap-labeled target RNA. The regions targeted by the dsRNAs are indicated as black bars on both sides. The 20-23 nt spacing between the predominant cleavage sites for the 111 bp long dsRNA is shown. The horizontal arrow indicates unspecific cleavage not due to RNAi. (B) Position of the cleavage sites on sense and antisense target RNAs. The sequences of the capped 177 nt sense and 180 nt antisense target RNAs are represented in antiparallel orientation such that complementary sequence are opposing each other. The region targeted by the different dsRNAs are indicated by differently colored bars positioned between sense and antisense target sequences. Cleavage sites are indicated by circles: large circle for strong cleavage, small circle for weak cleavage. The ^{32}P-radiolabeled phosphate group is marked by an asterisk.

Figure 4: 21 and 22 nt RNA fragments are generated by an RNase III-like mechanism.

(A) Sequences of ~21 nt RNAs after dsRNA processing. The ~21 nt RNA fragments generated by dsRNA processing were directionally cloned and sequenced. Oligoribonucleotides originating from the sense strand of the dsRNA are indicated as blue lines, those originating from the antisense strand as red lines. Thick bars are used if the same sequence was present in multiple clones, the number at the right indicating the frequency. The target RNA cleavage sites mediated by the dsRNA are indicated as orange circles, large circle for strong cleavage, small circle for weak cleavage (see Figure 3B). Circles on top of the sense strand indicated cleavage sites within the sense target and circles at the bottom of the dsRNA indicate cleavage site in the antisense target. Up to five additional nucleotides were identified in ~21 nt fragments derived from the 3' ends of the dsRNA. These nucleotides are random combinations of predominantly C, G, or A residues and were most likely added in an untemplated fashion during T7 transcription of the dsRNA-constituting strands. (B) Two-dimensional TLC analysis of the nucleotide composition of ~21 nt RNAs. The ~21 nt RNAs

were generated by incubation of internally radiolabeled 504 bp Pp-luc dsRNA in Drosophila lysate, gel-purified, and then digested to mononucleotides with nuclease P1 (top row) or ribonuclease T2 (bottom row). The dsRNA was internally radiolabeled by transcription in the presence of one of the indicated α-^{32}P nucleoside triphosphates. Radioactivity was detected by phosphorimaging. Nucleoside 5'-monophosphates, nucleoside 3'-monophosphates, nucleoside 5',3'-diphosphates, and inorganic phosphate are indicated as pN, Np, pNp, and p$_i$, respectively. Black circles indicate UV-absorbing spots from non-radioactive carrier nucleotides. The 3',5'-bisphosphates (red circles) were identified by co-migration with radiolabeled standards prepared by 5'-phosphorylation of nucleoside 3'-monophosphates with T4 polynucleotide kinase and γ-^{32}P-ATP.

Figure 5: Synthetic 21 and 22 nt RNAs Mediate Target RNA Cleavage.
(A) Graphic representation of control 52 bp dsRNA and synthetic 21 and 22 nt dsRNAs. The sense strand of 21 and 22 nt short interfering RNAs (siRNAs) is shown blue, the antisense strand in red. The sequences of the siRNAs were derived from the cloned fragments of 52 and 111 bp dsRNAs (Figure 4A), except for the 22 nt antisense strand of duplex 5. The siRNAs in duplex 6 and 7 were unique to the 111 bp dsRNA processing reaction. The two 3' overhanging nucleotides indicated in green are present in the sequence of the synthetic antisense strand of duplexes 1 and 3. Both strands of the control 52 bp dsRNA were prepared by in vitro transcription and a fraction of transcripts may contain untemplated 3' nucleotide addition. The target RNA cleavage sites directed by the siRNA duplexes are indicated as orange circles (see legend to Figure 4A) and were determined as shown in Figure 5B. (B) Position of the cleavage sites on sense and antisense target RNAs. The target RNA sequences are as described in Figure 3B. Control 52 bp dsRNA (10 nM) or 21 and 22 nt RNA duplexes 1-7 (100 nM) were incubated with target RNA for 2.5 h at 25°C in Drosophila lysate. The stable 5' cleavage products were resolved on the gel. The cleavage sites are indicated in Figure 5A. The region targeted by the 52 bp

dsRNA or the sense (s) or antisense (as) strands are indicated by the black bars to the side of the gel. The cleavage sites are all located within the region of identity of the dsRNAs. For precise determination of the cleavage sites of the antisense strand, a lower percentage gel was used.

Figure 6: Long 3' overhangs on short dsRNAs inhibit RNAi.

(A) Graphic representation of 52 bp dsRNA constructs. The 3' extensions of sense and antisense strand are indicated in blue and red, respectively. The observed cleavage sites on the target RNAs are represented as orange circles analogous to Figure 4A and were determined as shown in Figure 6B. (B) Position of the cleavage sites on sense and antisense target RNAs. The target RNA sequences are as described in Figure 3B. DsRNA (10 nM) was incubated with target RNA for 2.5 h at 25°C in Drosophila lysate. The stable 5' cleavage products were resolved on the gel. The major cleavage sites are indicated with a horizontal arrow and also represented in Figure 6A. The region targeted by the 52 bp dsRNA is represented as black bar at both sides of the gel.

Figure 7: Proposed Model for RNAi.

RNAi is predicted to begin with processing of dsRNA (sense strand in black, antisense strand in red) to predominantly 21 and 22 nt short interfering RNAs (siRNAs). Short overhanging 3' nucleotides, if present on the dsRNA, may be beneficial for processing of short dsRNAs. The dsRNA-processing proteins, which remain to be characterized, are represented as green and blue ovals, and assembled on the dsRNA in asymmetric fashion. In our model, this is illustrated by binding of a hypothetical blue protein or protein domain with the siRNA strand in 3' to 5' direction while the hypothetical green protein or protein domain is always bound to the opposing siRNA strand. These proteins or a subset remain associated with the siRNA duplex and preserve its orientation as determined by the direction of the dsRNA processing reaction. Only the siRNA sequence associated with the blue protein is able to guide target RNA cleavage. The endonuclease com-

- 18 -

plex is referred to as small interfering ribonucleoprotein complex or siRNP. It is presumed here, that the endonuclease that cleaves the dsRNA may also cleave the target RNA, probably by temporarily displacing the passive siRNA strand not used for target recognition. The target RNA is then cleaved in the center of the region recognized by the sequence-complementary guide siRNA.

Figure 8: Reporter constructs and siRNA duplexes.

(a) The firefly (Pp-luc) and sea pansy (Rr-luc) luciferase reporter gene regions from plasmids pGL2-Control, pGL-3-Control and pRL-TK (Promega) are illustrated. SV40 regulatory elements, the HSV thymidine kinase promoter and two introns (lines) are indicated. The sequence of GL3 luciferase is 95% identical to GL2, but RL is completely unrelated to both. Luciferase expression from pGL2 is approx. 10-fold lower than from pGL3 in transfected mammalian cells. The region targeted by the siRNA duplexes is indicated as black bar below the coding region of the luciferase genes. (b) The sense (top) and antisense (bottom) sequences of the siRNA duplexes targeting GL2, GL3 and RL luciferase are shown. The GL2 and GL3 siRNA duplexes differ by only 3 single nucleotide substitutions (boxed in gray). As unspecific control, a duplex with the inverted GL2 sequence, invGL2, was synthesized. The 2 nt 3' overhang of 2'-deoxythymidine is indicated as TT; uGL2 is similar to GL2 siRNA but contains ribo-uridine 3' overhangs.

Figure 9: RNA interference by siRNA duplexes.

Ratios of target control luciferase were normalized to a buffer control (bu, black bars); gray bars indicate ratios of *Photinus pyralis* (Pp-luc) GL2 or GL3 luciferase to *Renilla reniformis* (Rr-luc) RL luciferase (left axis), white bars indicate RL to GL2 or GL3 ratios (right axis). Panels a, c, e, g and i describe experiments performed with the combination of pGL2-Control and pRL-TK reporter plasmids, panels b, d, f, h and j with pGL3-Control and pRL-TK reporter plasmids. The cell line used for the interference experiment is indicated at the top of each plot. The ratios of Pp-luc/Rr-luc for the

- 19 -

buffer control (bu) varied between 0.5 and 10 for pGL2/pRL and between 0.03 and 1 for pGL3/pRL, respectively, before normalization and between the various cell lines tested. The plotted data were averaged from three independent experiments ± S.D.

Figure 10: Effects of 21 nt siRNA, 50 bp and 500 bp dsRNAs on luciferase expression in HeLa cells.

The exact length of the long dsRNAs is indicated below the bars. Panels a, c and e describe experiments performed with pGL2-Control and pRL-TK reporter plasmids, panels b, d and f with pGL3-Control and pRL-TK reporter plasmids. The data were averaged from two independent experiments ± S.D. (a), (b) Absolute Pp-luc expression, plotted in arbitrary luminescence units. (c), (d) Rr-luc expression, plotted in arbitrary luminescence units. (e), (f) Ratios of normalized target to control luciferase. The ratios of luciferase activity for siRNA duplexes were normalized to a buffer control (bu, black bars); the luminescence ratios for 50 or 500 bp dsRNAs were normalized to the respective ratios observed for 50 and 500 bp dsRNA from humanized GFP (hG, black bars). It should be noted that the overall differences in sequences between the 49 and 484 bp dsRNAs targeting GL2 and GL3 are not sufficient to confer specificity between GL2 and GL3 targets (43 nt uninterrupted identity in 49 bp segment, 239 nt longest uninterrupted identity in 484 bp segment).

Figure 11: Variation of the 3' overhang of duplexes of 21-nt siRNAs.

(A) Outline of the experimental strategy. The capped and polyadenylated sense target mRNA is depicted and the relative positions of sense and antisense siRNAs are shown. Eight series of duplexes, according to the eight different antisense strands were prepared. The siRNA sequences and the number of overhanging nucleotides were changed in 1-nt steps. (B) Normalized relative luminescence of target luciferase (Photinus pyralis, Pp-luc) to control luciferase (Renilla reniformis, Rr-luc) in D. melanogaster embryo lysate in the presence of 5 nM blunt-ended dsRNAs. The lumi-

nescence ratios determined in the presence of dsRNA were normalized to the ratio obtained for a buffer control (bu, black bar). Normalized ratios less than 1 indicate specific interference. (C-J) Normalized interference ratios for eight series of 21-nt siRNA duplexes. The sequences of siRNA duplexes are depicted above the bar graphs. Each panel shows the interference ratio for a set of duplexes formed with a given antisense guide siRNA and 5 different sense siRNAs. The number of overhanging nucleotides (3' over-hang, positive numbers; 5' overhangs, negative numbers) is indicated on the x-axis. Data points were averaged from at least 3 independent experiments, error bars represent standard deviations.

Figure 12: Variation of the length of the sense strand of siRNA duplexes. (A) Graphic representation of the experiment. Three 21-nt antisense strands were paired with eight sense siRNAs. The siRNAs were changed in length at their 3' end. The 3' overhang of the antisense siRNA was 1-nt (B), 2-nt (C), or 3-nt (D) while the sense siRNA overhang was varied for each series. The sequences of the siRNA duplexes and the corresponding interference ratios are indicated.

Figure 13: Variation of the length of siRNA duplexes with preserved 2-nt 3' overhangs.
(A) Graphic representation of the experiment. The 21-nt siRNA duplex is identical in sequence to the one shown in Figure 11H or 12C. The siRNA duplexes were extended to the 3' side of the sense siRNA (B) or the 5' side of the sense siRNA (C). The siRNA duplex sequences and the respective interference ratios are indicated.

Figure 14: Substitution of the 2'-hydroxyl groups of the siRNA ribose residues.
The 2'-hydroxyl groups (OH) in the strands of siRNA duplexes were replaced by 2'-deoxy (d) or 2'-O-methyl (Me). 2-nt and 4-nt 2'-deoxy substitu-

tions at the 3'-ends are indicated as 2-nt d and 4-nt d, respectively. Uridine residues were replaced by 2'-deoxy thymidine.

Figure 15: Mapping of sense and antisense target RNA cleavage by 21-nt siRNA duplexes with 2-nt 3' overhangs.

(A) Graphic representation of 32P (asterisk) cap-labelled sense and anti-sense target RNAs and siRNA duplexes. The position of sense and anti-sense target RNA cleavage is indicated by triangles on top and below the siRNA duplexes, respectively. (B) Mapping of target RNA cleavage sites. After 2 h incubation of 10 nM target with 100 nM siRNA duplex in D. melanogaster embryo lysate, the 5' cap-labelled substrate and the 5' cleavage products were resolved on sequencing gels. Length markers were generated by partial RNase T1 digestion (T1) and partial alkaline hydrolysis (OH-) of the target RNAs. The bold lines to the left of the images indicate the region covered by the siRNA strands 1 and 5 of the same orientation as the target.

Figure 16: The 5' end of a guide siRNA defines the position of target RNA cleavage.

(A, B) Graphic representation of the experimental strategy. The antisense siRNA was the same in all siRNA duplexes, but the sense strand was varied between 18 to 25 nt by changing the 3' end (A) or 18 to 23 nt by changing the 5' end (B). The position of sense and antisense target RNA cleavage is indicated by triangles on top and below the siRNA duplexes, respectively. (C, D) Analysis of target RNA cleavage using cap-labelled sense (top panel) or antisense (bottom panel) target RNAs. Only the cap-labelled 5' cleavage products are shown. The sequences of the siRNA duplexes are indicated, and the length of the sense siRNA strands is marked on top of the panel. The control lane marked with a dash in panel (C) shows target RNA incubated in absence of siRNAs. Markers were as described in Figure 15. The arrows in (D), bottom panel, indicate the target RNA cleavage sites that differ by 1 nt.

- 22 -

Figure 17: Sequence variation of the 3' overhang of siRNA duplexes.

The 2-nt 3' overhang (NN, in gray) was changed in sequence and composi-tion as indicated (T, 2'-deoxythymidine, dG, 2'-deoxyguanosine; asterisk, wild-type siRNA duplex). Normalized interference ratios were determined as described in Figure 11. The wild-type sequence is the same as depicted in Figure 14.

Figure 18: Sequence specificity of target recognition.

The sequences of the mismatched siRNA duplexes are shown, modified sequence segments or single nucleotides are underlayed in gray. The reference duplex (ref) and the siRNA duplexes 1 to 7 contain 2'-deoxythy-midine 2-nt overhangs. The silencing efficiency of the thymidine-modified reference duplex was comparable to the wild-type sequence (Figure 17). Normalized interference ratios were determined as described in Figure 11.

Figure 19: Variation of the length of siRNA duplexes with preserved 2-nt 3' overhangs.

The siRNA duplexes were extended to the 3' side of the sense siRNA (A) or the 5' side of the sense siRNA (B). The siRNA duplex sequences and the respective interference ratios are indicated. For HeLa SS6 cells, siRNA duplexes (0.84 μg) targeting GL2 luciferase were transfected together with pGL2-Control and pRL-TK plasmids. For comparison, the in vitro RNAi activities of siRNA duplexes tested in D. melanogaster lysate are indicated.

Example 1

RNA Interference Mediated by Small Synthetic RNAs

1.1.   Experimental Procedures

1.1.1 In Vitro RNAi

In vitro RNAi and lysate preparations were performed as described previously (Tuschl et al., 1999; Zamore et al., 2000). It is critical to use

freshly dissolved creatine kinase (Roche) for optimal ATP regeneration. The RNAi translation assays (Fig. 1) were performed with dsRNA concentrations of 5 nM and an extended pre-incubation period of 15 min at 25°C prior to the addition of in vitro transcribed, capped and polyadenylated Pp-luc and Rr-luc reporter mRNAs. The incubation was continued for 1 h and the relative amount of Pp-luc and Rr-luc protein was analyzed using the dual luciferase assay (Promega) and a Monolight 3010C luminometer (PharMingen).

1.1.2 RNA Synthesis

Standard procedures were used for in vitro transcription of RNA from PCR templates carrying T7 or SP6 promoter sequences, see for example (Tuschl et al., 1998). Synthetic RNA was prepared using Expedite RNA phosphoramidites (Proligo). The 3′ adapter oligonucleotide was synthesized using dimethoxytrityl-1,4-benzenedimethanol-succinyl-aminopropyl-CPG. The oligoribonucleotides were deprotected in 3 ml of 32% ammonia/ethanol (3/1) for 4 h at 55°C (Expedite RNA) or 16 h at 55°C (3′ and 5′ adapter DNA/RNA chimeric oligonucleotides) and then desilylated and gel-purified as described previously (Tuschl et al., 1993). RNA transcripts for dsRNA preparation including long 3′ overhangs were generated from PCR templates that contained a T7 promoter in sense and an SP6 promoter in antisense direction. The transcription template for sense and antisense target RNA was PCR-amplified with GCG<u>TAATACGACTCACTATA</u>GAACAATTGCTTTTACAG (underlined, T7 promoter) as 5′ primer and <u>ATTTAGGTGACACTATA</u>GGCATAAAGAATTGAAGA (underlined, SP6 promoter) as 3′ primer and the linearized Pp-luc plasmid (pGEM-luc sequence) (Tuschl et al., 1999) as template; the T7-transcribed sense RNA was 177 nt long with the Pp-luc sequence between pos. 113-273 relative to the start codon and followed by 17 nt of the complement of the SP6 promoter sequence at the 3′ end. Transcripts for blunt-ended dsRNA

- 24 -

formation were prepared by transcription from two different PCR products which only contained a single promoter sequence.

DsRNA annealing was carried out using a phenol/chloroform extraction. Equimolar concentration of sense and antisense RNA (50 nM to 10 μM, depending on the length and amount available) in 0.3 M NaOAc (pH 6) were incubated for 30 s at 90°C and then extracted at room temperature with an equal volume of phenol/chloroform, and followed by a chloroform extraction to remove residual phenol. The resulting dsRNA was precipitated by addition of 2.5-3 volumes of ethanol. The pellet was dissolved in lysis buffer (100 mM KCl, 30 mM HEPES-KOH, pH 7.4, 2 mM Mg(OAc)$_2$) and the quality of the dsRNA was verified by standard agarose gel electro-phoreses in 1 x TAE-buffer. The 52 bp dsRNAs with the 17 nt and 20 nt 3' overhangs (Figure 6) were annealed by incubating for 1 min at 95 °C, then rapidly cooled to 70°C and followed by slow cooling to room temperature over a 3 h period (50 μl annealing reaction, 1 μM strand concentration, 300 mM NaCl, 10 mM Tris-HCl, pH 7.5). The dsRNAs were then phenol/ chloroform extracted, ethanol-precipitated and dissolved in lysis buffer.

Transcription of internally ^{32}P-radiolabeled RNA used for dsRNA preparation (Figures 2 and 4) was performed using 1 mM ATP, CTP, GTP, 0.1 or 0.2 mM UTP, and 0.2-0.3 μM -^{32}P-UTP (3000 Ci/mmol), or the respective ratio for radiolabeled nucleoside triphosphates other than UTP. Labeling of the cap of the target RNAs was performed as described previously. The target RNAs were gel-purified after cap-labeling.

1.1.3 Cleavage Site Mapping

Standard RNAi reactions were performed by pre-incubating 10 nM dsRNA for 15 min followed by addition of 10 nM cap-labeled target RNA. The reaction was stopped after a further 2 h (Figure 2A) or 2.5 h incubation (Figure 5B and 6B) by proteinase K treatment (Tuschl et al., 1999). The samples were then analyzed on 8 or 10% sequencing gels. The 21 and 22

nt synthetic RNA duplexes were used at 100 nM final concentration (Fig 5B).

1.1.4 Cloning of ~21 nt RNAs

The 21 nt RNAs were produced by incubation of radiolabeled dsRNA in Drosophila lysate in absence of target RNA (200 μl reaction, 1 h incubation, 50 nM dsP111, or 100 nM dsP52 or dsP39). The reaction mixture was subsequently treated with proteinase K (Tuschl et al., 1999) and the dsRNA-processing products were separated on a denaturing 15% poly-acrylamide gel. A band, including a size range of at least 18 to 24 nt, was excised, eluted into 0.3 M NaCl overnight at 4°C and in siliconized tubes. The RNA was recovered by ethanol-precipitation and dephosphorylated (30 μl reaction, 30 min, 50°C, 10 U alkaline phosphatase, Roche). The reaction was stopped by phenol/chloroform extraction and the RNA was ethanol-precipitated. The 3' adapter oligonucleotide (pUUUaaccgcatccttctcx: uppercase, RNA; lowercase, DNA; p, phosphate; x, 4-hydroxymethylbenzyl) was then ligated to the dephosphorylated ~21 nt RNA (20 μl reaction, 30 min, 37°C, 5 μM 3' adapter, 50 mM Tris-HCl, pH 7.6, 10 mM MgCl$_2$, 0.2 mM ATP, 0.1 mg/ml acetylated BSA, 15% DMSO, 25 U T4 RNA ligase, Amersham-Pharmacia) (Pan and Uhlenbeck, 1992). The ligation reaction was stopped by the addition of an equal volume of 8 M urea/50 mM EDTA stopmix and directly loaded on a 15% gel. Ligation yields were greater 50%. The ligation product was recovered from the gel and 5'-phosphorylated (20 μl reaction, 30 min, 37°C, 2 mM ATP, 5 U T4 polynucleotide kinase, NEB). The phosphorylation reaction was stopped by phenol/chloroform extraction and RNA was recovered by ethanol-precipitation. Next, the 5' adapter (tactaatacgactcactAAA: uppercase, RNA; lowercase, DNA) was ligated to the phosphorylated ligation product as described above. The new ligation product was gel-purified and eluted from the gel slice in the presence of reverse transcription primer (GACTAGCTG**GAATTC**AAGGATGCGGTTAAA: bold, Eco RI site) used as carrier. Reverse transcription (15 μl reaction, 30 min, 42°C, 150 U Super-

script II reverse transcriptase, Life Technologies) was followed by PCR using as 5' primer CAGCCAACG**GAATTC**ATACGACTCACTAAA (bold, Eco RI site) and the 3' RT primer. The PCR product was purified by phenol/ chloroform extraction and ethanol-precipitated. The PCR product was then digested with Eco RI (NEB) and concatamerized using T4 DNA ligase (high conc., NEB). Concatamers of a size range of 200 to 800 bp were separated on a low-melt agarose gel, recovered from the gel by a standard melting and phenol extraction procedure, and ethanol-precipitated. The unpaired ends were filled in by incubation with Taq polymerase under standard conditions for 15 min at 72°C and the DNA product was directly ligated into the pCR2.1-TOPO vector using the TOPO TA cloning kit (Invitrogen). Colonies were screened using PCR and M13-20 and M13 Reverse sequencing primers. PCR products were directly submitted for custom sequencing (Sequence Laboratories Göttingen GmbH, Germany). On average, four to five 21mer sequences were obtained per clone.

1.1.5 2D-TLC Analysis

Nuclease P1 digestion of radiolabeled, gel-purified siRNAs and 2D-TLC was carried out as described (Zamore et al., 2000). Nuclease T2 digestion was performed in 10 μl reactions for 3 h at 50°C in 10 mM ammonium acetate (pH 4.5) using 2 μg/μl carrier tRNA and 30 U ribonuclease T2 (Life Technologies). The migration of non-radioactive standards was determined by UV shadowing. The identity of nucleoside-3',5'-disphosphates was confirmed by co-migration of the T2 digestion products with standards prepared by 5'-^{32}P-phosphorylation of commercial nucleoside 3'-monophosphates using γ-32P-ATP and T4 polynucleotide kinase (data not shown).

1.2 Results and Discussion

1.2.1 Length Requirements for Processing of dsRNA to 21 and 22 nt RNA Fragments

5   Lysate prepared from D. melanogaster syncytial embryos recapitulates RNAi in vitro providing a novel tool for biochemical analysis of the mechanism of RNAi (Tuschl et al., 1999; Zamore et al., 2000). In vitro and in vivo analysis of the length requirements of dsRNA for RNAi has revealed that short dsRNA (<150 bp) are less effective than longer dsRNAs in

10 degrading target mRNA (Caplen et al., 2000; Hammond et al., 2000; Ngo et al., 1998); Tuschl et al., 1999). The reasons for reduction in mRNA degrading efficiency are not understood. We therefore examined the precise length requirement of dsRNA for target RNA degradation under optimized conditions in the Drosophila lysate (Zamore et al., 2000). Several

15 series of dsRNAs were synthesized and directed against firefly luciferase (Pp-luc) reporter RNA. The specific suppression of target RNA expression was monitored by the dual luciferase assay (Tuschl et al., 1999) (Figures 1A and 1B). We detected specific inhibition of target RNA expression for dsRNAs as short as 38 bp, but dsRNAs of 29 to 36 bp were not effective

20 in this process. The effect was independent of the target position and the degree of inhibition of Pp-luc mRNA expression correlated with the length of the dsRNA, i.e. long dsRNAs were more effective than short dsRNAs.

It has been suggested that the 21-23 nt RNA fragments generated by

25 processing of dsRNAs are the mediators of RNA interference and co-suppression (Hamilton and Baulcombe, 1999; Hammond et al., 2000; Zamore et al., 2000). We therefore analyzed the rate of 21-23 nt fragment formation for a subset of dsRNAs ranging in size between 501 to 29 bp. Formation of 21-23 nt fragments in Drosophila lysate (Figure 2) was readily

30 detectable for 39 to 501 bp long dsRNAs but was significantly delayed for the 29 bp dsRNA. This observation is consistent with a role of 21-23 nt fragments in guiding mRNA cleavage and provides an explanation for the

lack of RNAi by 30 bp dsRNAs. The length dependence of 21-23 mer formation is likely to reflect a biologically relevant control mechanism to prevent the undesired activation of RNAi by short intramolecular base-paired structures of regular cellular RNAs.

1.2.2 39 bp dsRNA Mediates Target RNA Cleavage at a Single Site

Addition of dsRNA and 5'-capped target RNA to the Drosophila lysate results in sequence-specific degradation of the target RNA (Tuschl et al., 1999). The target mRNA is only cleaved within the region of identity with the dsRNA and many of the target cleavage sites were separated by 21-23 nt (Zamore et al., 2000). Thus, the number of cleavage sites for a given dsRNA was expected to roughly correspond to the length of the dsRNA divided by 21. We mapped the target cleavage sites on a sense and an antisense target RNA which was 5' radiolabeled at the cap (Zamore et al., 2000) (Figures 3A and 3B). Stable 5' cleavage products were separated on a sequencing gel and the position of cleavage was determined by comparison with a partial RNase T1 and an alkaline hydrolysis ladder from the target RNA.

Consistent with the previous observation (Zamore et al., 2000), all target RNA cleavage sites were located within the region of identity to the dsRNA. The sense or the antisense traget was only cleaved once by 39 bp dsRNA. Each cleavage site was located 10 nt from the 5' end of the region covered by the dsRNA (Figure 3B). The 52 bp dsRNA, which shares the same 5' end with the 39 bp dsRNA, produces the same cleavage site on the sense target, located 10 nt from the 5' end of the region of identity with the dsRNA, in addition to two weaker cleavage sites 23 and 24 nt downstream of the first site. The antisense target was only cleaved once, again 10 nt from the 5' end of the region covered by its respective dsRNA. Mapping of the cleavage sites for the 38 to 49 bp dsRNAs shown in Figure 1 showed that the first and predominant cleavage site was always located 7 to 10 nt downstream of the region covered by the dsRNA (data not

shown). This suggests that the point of target RNA cleavage is determined by the end of the dsRNA and could imply that processing to 21-23 mers starts from the ends of the duplex.

Cleavage sites on sense and antisense target for the longer 111 bp dsRNA were much more frequent than anticipated and most of them appear in clusters separated by 20 to 23 nt (Figures 3A and 3B). As for the shorter dsRNAs, the first cleavage site on the sense target is 10 nt from the 5' end of the region spanned by the dsRNA; and the first cleavage site on the antisense target is located 9 nt from the 5' end of region covered by the dsRNA. It is unclear what causes this disordered cleavage, but one possibility could be that longer dsRNAs may not only get processed from the ends but also internally, or there are some specificity determinants for dsRNA processing which we do not yet understand. Some irregularities to the 21-23 nt spacing were also previously noted (Zamore et al., 2000). To better understand the molecular basis of dsRNA processing and target RNA recognition, we decided to analyze the sequences of the 21-23 nt fragments generated by processing of 39, 52, and 111 bp dsRNAs in the Drosophila lysate.

1.2.3 dsRNA is Processed to 21 and 22 nt RNAs by an RNase III-Like Mechanism

In order to characterize the 21-23 nt RNA fragments we examined the 5' and 3' termini of the RNA fragments. Periodate oxidation of gel-purified 21-23 nt RNAs followed by ß-elimination indicated the presence of a terminal 2' and 3' hydroxyl groups. The 21-23 mers were also responsive to alkaline phosphatase treatment indicating the presence of a 5' terminal phosphate group. The presence of 5' phosphate and 3' hydroxyl termini suggests that the dsRNA could be processed by an enzymatic activity similar to E. coli RNase III (for reviews, see (Dunn, 1982; Nicholson, 1999; Robertson, 1990; Robertson, 1982)).

Directional cloning of 21-23 nt RNA fragments was performed by ligation of a 3' and 5' adapter oligonucleotide to the purified 21-23 mers using T4 RNA ligase. The ligation products were reverse transcribed, PCR-amplified, concatamerized, cloned, and sequenced. Over 220 short RNAs were sequenced from dsRNA processing reactions of the 39, 52 and 111 bp dsRNAs (Figure 4A). We found the following length distribution: 1% 18 nt, 5% 19 nt, 12% 20 nt, 45% 21 nt, 28% 22 nt, 6% 23 nt, and 2% 24 nt. Sequence analysis of the 5' terminal nucleotide of the processed fragments indicated that oligonucleotides with a 5' guanosine were underrepresented. This bias was most likely introduced by T4 RNA ligase which discriminates against 5' phosphorylated guanosine as donor oligonucleotide; no significant sequence bias was seen at the 3' end. Many of the ~21 nt fragments derived from the 3' ends of the sense or antisense strand of the duplexes include 3' nucleotides that are derived from untemplated addition of nucleotides during RNA synthesis using T7 RNA polymerase. Interestingly, a significant number of endogenous Drosophila ~21 nt RNAs were also cloned, some of them from LTR and non-LTR retrotransposons (data not shown). This is consistent with a possible role for RNAi in transposon silencing.

The ~21 nt RNAs appear in clustered groups (Figure 4A) which cover the entire dsRNA sequences. Apparently, the processing reaction cuts the dsRNA by leaving staggered 3' ends, another characteristic of RNase III cleavage. For the 39 bp dsRNA, two clusters of ~21 nt RNAs were found from each dsRNA-constituting strand including overhanging 3' ends, yet only one cleavage site was detected on the sense and antisense target (Figures 3A and 3B). If the ~21 nt fragments were present as single-stranded guide RNAs in a complex that mediates mRNA degradation, it could be assumed that at least two target cleavage sites exist, but this was not the case. This suggests that the ~21 nt RNAs may be present in double-stranded form in the endonuclease complex but that only one of the strands can be used for target RNA recognition and cleavage. The use of

only one of the ~21 nt strands for target cleavage may simply be deter-
mined by the orientation in which the ~21 nt duplex is bound to the nucle-
ase complex. This orientation is defined by the direction in which the
original dsRNA was processed.

The ~21mer clusters for the 52 bp and 111 bp dsRNA are less well de-
fined when compared to the 39 bp dsRNA. The clusters are spread over
regions of 25 to 30 nt most likely representing several distinct subpopula-
tions of ~21 nt duplexes and therefore guiding target cleavage at several
nearby sites. These cleavage regions are still predominantly separated by
20 to 23 nt intervals. The rules determining how regular dsRNA can be
processed to ~21 nt fragments are not yet understood, but it was previ-
ously observed that the approx. 21-23 nt spacing of cleavage sites could
be altered by a run of uridines (Zamore et al., 2000). The specificity of
dsRNA cleavage by E. coli RNase III appears to be mainly controlled by
antideterminants, i.e. excluding some specific base-pairs at given positions
relative to the cleavage site (Zhang and Nicholson, 1997).

To test whether sugar-, base- or cap-modification were present in
processed ~21 nt RNA fragments, we incubated radiolabeled 505 bp Pp-
luc dsRNA in lysate for 1 h, isolated the ~21 nt products, and digested it
with P1 or T2 nuclease to mononucleotides. The nucleotide mixture was
then analyzed by 2D thin-layer chromatography (Figure 4B). None of the
four natural ribonucleotides were modified as indicated by P1 or T2 di-
gestion. We have previously analyzed adenosine to inosine conversion in
the ~21 nt fragments (after a 2 h incubation) and detected a small extent
(<0.7%) deamination (Zamore et al., 2000); shorter incubation in lysate
(1 h) reduced this inosine fraction to barely detectable levels. RNase T2,
which cleaves 3' of the phosphodiester linkage, produced nucleoside 3'-
phosphate and nucleoside 3',5'-diphosphate, thereby indicating the pre-
sence of a 5'-terminal monophosphate. All four nucleoside 3',5'-diphos-
phates were detected and suggest that the internucleotidic linkage was

cleaved with little or no sequence-specificity. In summary, the ~21 nt fragments are unmodified and were generated from dsRNA such that 5'-monophosphates and 3'-hydroxyls were present at the 5'-end.

1.2.4 Synthetic 21 and 22 nt RNAs Mediate Target RNA Cleavage

Analysis of the products of dsRNA processing indicated that the ~21 nt fragments are generated by a reaction with all the characteristics of an RNase III cleavage reaction (Dunn, 1982; Nicholson, 1999; Robertson, 1990; Robertson, 1982). RNase III makes two staggered cuts in both strands of the dsRNA, leaving a 3' overhang of about 2 nt. We chemically synthesized 21 and 22 nt RNAs, identical in sequence to some of the cloned ~21 nt fragments, and tested them for their ability to mediate target RNA degradation (Figures 5A and 5B). The 21 and 22 nt RNA duplexes were incubated at 100 nM concentrations in the lysate, a 10-fold higher concentrations than the 52 bp control dsRNA. Under these conditions, target RNA cleavage is readily detectable. Reducing the concentration of 21 and 22 nt duplexes from 100 to 10 nM does still cause target RNA cleavage. Increasing the duplex concentration from 100 nM to 1000 nM however does not further increase target cleavage, probably due to a limiting protein factor within the lysate.

In contrast to 29 or 30 bp dsRNAs that did not mediate RNAi, the 21 and 22 nt dsRNAs with overhanging 3' ends of 2 to 4 nt mediated efficient degradation of target RNA (duplexes 1, 3, 4, 6, Figures 5A and 5B). Blunt-ended 21 or 22 nt dsRNAs (duplexes 2, 5, and 7, Figures 5A and 5B) were reduced in their ability to degrade the target and indicate that overhanging 3' ends are critical for reconstitution of the RNA-protein nuclease complex. The single-stranded overhangs may be required for high affinity binding of the ~ 21 nt duplex to the protein components. A 5' terminal phosphate, although present after dsRNA processing, was not required to mediate target RNA cleavage and was absent from the short synthetic RNAs.

The synthetic 21 and 22 nt duplexes guided cleavage of sense as well as antisense targets within the region covered by the short duplex. This is an important result considering that a 39 bp dsRNA, which forms two pairs of clusters of ~21 nt fragments (Fig. 2), cleaved sense or antisense target only once and not twice. We interpret this result by suggesting that only one of two strands present in the ~21 nt duplex is able to guide target RNA cleavage and that the orientation of the ~21 nt duplex in the nuclease complex is determined by the initial direction of dsRNA processing. The presentation of an already perfectly processed ~21 nt duplex to the in vitro system however does allow formation of the active sequence-specific nuclease complex with two possible orientations of the symmetric RNA duplex. This results in cleavage of sense as well as antisense target within the region of identity with the 21 nt RNA duplex.

The target cleavage site is located 11 or 12 nt downstream of the first nucleotide that is complementary to the 21 or 22 nt guide sequence, i.e. the cleavage site is near center of the region covered by the 21 or 22 nt RNAs (Figures 4A and 4B). Displacing the sense strand of a 22 nt duplex by two nucleotides (compare duplexes 1 and 3 in Figure 5A) displaced the cleavage site of only the antisense target by two nucleotides. Displacing both sense and antisense strand by two nucleotides shifted both cleavage sites by two nucleotides (compare duplexes 1 and 4). We predict that it will be possible to design a pair of 21 or 22 nt RNAs to cleave a target RNA at almost any given position.

The specificity of target RNA cleavage guided by 21 and 22 nt RNAs appears exquisite as no aberrant cleavage sites are detected (Figure 5B). It should however be noted, that the nucleotides present in the 3' overhang of the 21 and 22 nt RNA duplex may contribute less to substrate recognition than the nucleotides near the cleavage site. This is based on the observation that the 3' most nucleotide in the 3' overhang of the active duplexes 1 or 3 (Figure 5A) is not complementary to the target. A detailed

analysis of the specificity of RNAi can now be readily undertaken using synthetic 21 and 22 nt RNAs.

Based on the evidence that synthetic 21 and 22 nt RNAs with overhanging 3' ends mediate RNA interference, we propose to name the ~21 nt RNAs "short interfering RNAs" or siRNAs and the respective RNA-protein complex a "small interfering ribonucleoprotein particle" or siRNP.

1.2.5 3' Overhangs of 20 nt on short dsRNAs inhibit RNAi

We have shown that short blunt-ended dsRNAs appear to be processed from the ends of the dsRNA. During our study of the length dependence of dsRNA in RNAi, we have also analyzed dsRNAs with 17 to 20 nt overhanging 3' ends and found to our surprise that they were less potent than blunt-ended dsRNAs. The inhibitory effect of long 3' ends was particularly pronounced for dsRNAs up to 100 bp but was less dramatic for longer dsRNAs. The effect was not due to imperfect dsRNA formation based on native gel analysis (data not shown). We tested if the inhibitory effect of long overhanging 3' ends could be used as a tool to direct dsRNA processing to only one of the two ends of a short RNA duplex.

We synthesized four combinations of the 52 bp model dsRNA, blunt-ended, 3' extension on only the sense strand, 3'-extension on only the antisense strand, and double 3' extension on both strands, and mapped the target RNA cleavage sites after incubation in lysate (Figures 6A and 6B). The first and predominant cleavage site of the sense target was lost when the 3' end of the antisense strand of the duplex was extended, and vice versa, the strong cleavage site of the antisense target was lost when the 3' end of sense strand of the duplex was extended. 3' Extensions on both strands rendered the 52 bp dsRNA virtually inactive. One explanation for the dsRNA inactivation by ~20 nt 3' extensions could be the association of single-stranded RNA-binding proteins which could interfere with the association of one of the dsRNA-processing factors at this end. This result

is also consistent with our model where only one of the strands of the siRNA duplex in the assembled siRNP is able to guide target RNA cleavage. The orientation of the strand that guides RNA cleavage is defined by the direction of the dsRNA processing reaction. It is likely that the presence of 3' staggered ends may facilitate the assembly of the processing complex. A block at the 3' end of the sense strand will only permit dsRNA processing from the opposing 3' end of the antisense strand. This in turn generates siRNP complexes in which only the antisense strand of the siRNA duplex is able to guide sense target RNA cleavage. The same is true for the reciprocal situation.

The less pronounced inhibitory effect of long 3' extensions in the case of longer dsRNAs ($\geq$500 bp, data not shown) suggests to us that long dsRNAs may also contain internal dsRNA-processing signals or may get processed cooperatively due to the association of multiple cleavage factors.

1.2.6 A Model for dsRNA-Directed mRNA Cleavage

The new biochemical data update the model for how dsRNA targets mRNA for destruction (Figure 7). Double-stranded RNA is first processed to short RNA duplexes of predominantly 21 and 22 nt in length and with staggered 3' ends similar to an RNase III-like reaction (Dunn, 1982; Nicholson, 1999; Robertson, 1982). Based on the 21-23 nt length of the processed RNA fragments it has already been speculated that an RNase III-like activity may be involved in RNAi (Bass, 2000). This hypothesis is further supported by the presence of 5' phosphates and 3' hydroxyls at the termini of the siRNAs as observed in RNase III reaction products (Dunn, 1982; Nicholson, 1999). Bacterial RNase III and the eukaryotic homologs Rnt1p in S. cerevisiae and Pac1p in S. pombe have been shown to function in processing of ribosomal RNA as well as snRNA and snoRNAs (see for example Chanfreau et al., 2000).

- 36 -

Little is known about the biochemistry of RNase III homologs from plants, animals or human. Two families of RNase III enzymes have been identified predominantly by database-guided sequence analysis or cloning of cDNAs. The first RNase III family is represented by the 1327 amino acid long D. melanogaster protein drosha (Acc. AF116572). The C-terminus is composed of two RNase III and one dsRNA-binding domain and the N-terminus is of unknown function. Close homologs are also found in C. elegans (Acc. AF160248) and human (Acc. AF189011) (Filippov et al., 2000; Wu et al., 2000). The drosha-like human RNase III was recently cloned and characterized (Wu et al., 2000). The gene is ubiquitously expressed in human tissues and cell lines, and the protein is localized in the nucleus and the nucleolus of the cell. Based on results inferred from antisense inhibition studies, a role of this protein for rRNA processing was suggested. The second class is represented by the C. elegans gene K12H4.8 (Acc. S44849) coding for a 1822 amino acid long protein. This protein has an N-terminal RNA helicase motif which is followed by 2 RNase III catalytic domains and a dsRNA-binding motif, similar to the drosha RNase III family. There are close homologs in S. pombe (Acc. Q09884), A. thaliana (Acc. AF187317), D. melanogaster (Acc. AE003740), and human (Acc. AB028449) (Filippov et al., 2000; Jacobsen et al., 1999; Matsuda et al., 2000). Possibly the K12H4.8 RNase III/helicase is the likely candidate to be involved in RNAi.

Genetic screens in C. elegans identified rde-1 and rde-4 as essential for activation of RNAi without an effect on transposon mobilization or co-suppression (Dernburg et al., 2000; Grishok et al., 2000; Ketting and Plasterk, 2000; Tabara et al., 1999). This led to the hypothesis that these genes are important for dsRNA processing but are not involved in mRNA target degradation. The function of both genes is as yet unknown, the rde-1 gene product is a member of a family of proteins similar to the rabbit protein eIF2C (Tabara et al., 1999), and the sequence of rde-4 has not yet

- 37 -

been described. Future biochemical characterization of these proteins should reveal their molecular function.

Processing to the siRNA duplexes appears to start from the ends of both blunt-ended dsRNAs or dsRNAs with short (1-5 nt) 3' overhangs, and proceeds in approximately 21-23 nt steps. Long (~20 nt) 3' staggered ends on short dsRNAs suppress RNAi, possibly through interaction with single-stranded RNA-binding proteins. The suppression of RNAi by single-stranded regions flanking short dsRNA and the lack of siRNA formation from short 30 bp dsRNAs may explain why structured regions frequently encountered in mRNAs do not lead to activation of RNAi.

Without wishing to be bound by theory, we presume that the dsRNA-processing proteins or a subset of these remain associated with the siRNA duplex after the processing reaction. The orientation of the siRNA duplex relative to these proteins determines which of the two complementary strands functions in guiding target RNA degradation. Chemically synthesized siRNA duplexes guide cleavage of sense as well as antisense target RNA as they are able to associate with the protein components in either of the two possible orientation.

The remarkable finding that synthetic 21 and 22 nt siRNA duplexes can be used for efficient mRNA degradation provides new tools for sequence-specific regulation of gene expression in functional genomics as well as biomedical studies. The siRNAs may be effective in mammalian systems where long dsRNAs cannot be used due to the activation of the PKR response (Clemens, 1997). As such, the siRNA duplexes represent a new alternative to antisense or ribozyme therapeutics.

Example 2
RNA Interference in Human Tissue Cultures

2.1   Methods

2.1.1 RNA preparation

21 nt RNAs were chemically synthesized using Expedite RNA phosphorami-
dites and thymidine phosphoramidite (Proligo, Germany). Synthetic oligonu-
cleotides were deprotected and gel-purified (Example 1), followed by Sep-
Pak C18 cartridge (Waters, Milford, MA, USA) purification (Tuschl, 1993).
The siRNA sequences targeting GL2 (Acc. X65324) and GL3 luciferase
(Acc. U47296) corresponded to the coding regions 153-173 relative to the
first nucleotide of the start codon, siRNAs targeting RL (Acc. AF025846)
corresponded to region 119-129 after the start codon. Longer RNAs were
transcribed with T7 RNA polymerase from PCR products, followed by gel
and Sep-Pak purification. The 49 and 484 bp GL2 or GL3 dsRNAs corre-
sponded to position 113-161 and 113-596, respectively, relative to the
start of translation; the 50 and 501 bp RL dsRNAs corresponded to posi-
tion 118-167 and 118-618, respectively. PCR templates for dsRNA syn-
thesis targeting humanized GFP (hG) were amplified from pAD3 (Kehlen-
bach, 1998), whereby 50 and 501 bp hG dsRNA corresponded to position
118-167 and 118-618, respectively, to the start codon.

For annealing of siRNAs, 20 μM single strands were incubated in annealing
buffer (100 mM potassium acetate, 30 mM HEPES-KOH at pH 7.4, 2 mM
magnesium acetate) for 1 min at 90°C followed by 1 h at 37°C. The 37°C
incubation step was extended overnight for the 50 and 500 bp dsRNAs
and these annealing reactions were performed at 8.4 μM and 0.84 μM
strand concentrations, respectively.

2.1.2 Cell Culture

S2 cells were propagated in Schneider's *Drosophila* medium (Life Technologies) supplemented with 10% FBS, 100 units/ml penicillin and 100 μg/ml streptomycin at 25°C. 293, NIH/3T3, HeLa S3, COS-7 cells were grown at 37°C in Dulbecco's modified Eagle's medium supplemented with 10% FBS, 100 units/ml penicillin and 100 μg/ml streptomycin. Cells were regularly passaged to maintain exponential growth. 24 h before transfection at approx. 80% confluency, mammalian cells were trypsinized and diluted 1:5 with fresh medium without antibiotics (1-3 x 10^5 cells/ml) and transferred to 24-well plates (500 μl/well). S2 cells were not trypsinized before splitting. Transfection was carried out with Lipofectamine 2000 reagent (Life Technologies) as described by the manufacturer for adherent cell lines. Per well, 1.0 μg pGL2-Control (Promega) or pGL3-Control (Promega), 0.1 μg pRL-TK (Promega) and 0.28 μg siRNA duplex or dsRNA, formulated into liposomes, were applied; the final volume was 600 μl per well. Cells were incubated 20 h after transfection and appeared healthy thereafter. Luciferase expression was subsequently monitored with the Dual luciferase assay (Promega). Transfection efficiencies were determined by fluorescence microscopy for mammalian cell lines after co-transfection of 1.1 μg hGFP-encoding pAD3 and 0.28 μg invGL2 inGL2 siRNA and were 70-90%. Reporter plasmids were amplified in XL-1 Blue (Stratagene) and purified using the Qiagen EndoFree Maxi Plasmid Kit.

2.2   Results and Discussion

To test whether siRNAs are also capable of mediating RNAi in tissue culture, we synthesized 21 nt siRNA duplexes with symmetric 2 nt 3' overhangs directed against reporter genes coding for sea pansy (*Renilla reniformis*) and two sequence variants of firefly (*Photinus pyralis*, GL2 and GL3) luciferases (Fig. 8a, b). The siRNA duplexes were co-transfected with the reporter plasmid combinations pGL2/pRL or pGL3/pRL into *D. melanogaster Schneider S2* cells or mammalian cells using cationic liposomes. Luciferase activities were determined 20 h after transfection. In all cell lines tested,

we observed specific reduction of the expression of the reporter genes in the presence of cognate siRNA duplexes (Fig. 9a-j). Remarkably, the absolute luciferase expression levels were unaffected by non-cognate siRNAs, indicating the absence of harmful side effects by 21 nt RNA duplexes (e.g. Fig. 10a-d for HeLa cells). In *D. melanogaster* S2 cells (Fig. 9a, b), the specific inhibition of luciferases was complete. In mammalian cells, where the reporter genes were 50- to 100-fold stronger expressed, the specific suppression was less complete (Fig. 9c-j). GL2 expression was reduced 3- to 12-fold, GL3 expression 9- to 25-fold and RL expression 1- to 3-fold, in response to the cognate siRNAs. For 293 cells, targeting of RL luciferase by RL siRNAs was ineffective, although GL2 and GL3 targets responded specifically (Fig. 9i, j). The lack of reduction of RL expression in 293 cells may be due to its 5- to 20-fold higher expression compared to any other mammalian cell line tested and/or to limited accessibility of the target sequence due to RNA secondary structure or associated proteins. Nevertheless, specific targeting of GL2 and GL3 luciferase by the cognate siRNA duplexes indicated that RNAi is also functioning in 293 cells.

The 2 nt 3' overhang in all siRNA duplexes, except for uGL2, was composed of (2'-deoxy) thymidine. Substituion of uridine by thymidine in the 3' overhang was well tolerated in the *D. melanogaster* in vitro sytem and the sequence of the overhang was uncritical for target recognition. The thymidine overhang was chosen, because it is supposed to enhance nuclease resistance of siRNAs in the tissue culture medium and within transfected cells. Indeed, the thymidine-modified GL2 siRNA was slightly more potent than the unmodified uGL2 siRNA in all cell lines tested (Fig. 9a, c, e, g, i). It is conceivable that further modifications of the 3' overhanging nucleotides may provide additional benefits to the delivery and stability of siRNA duplexes.

In co-transfection experiments, 25 nM siRNA duplexes with respect to the final volume of tissue culture medium were used (Fig. 9, 10). Increasing

the siRNA concentration to 100 nM did not enhance the specific silencing effects, but started to affect transfection efficiencies due to competition for liposome encapsulation between plasmid DNA and siRNA (data not shown). Decreasing the siRNA concentration to 1.5 nM did not reduce the

5   specific silencing effect (data not shown), even though the siRNAs were now only 2- to 20-fold more concentrated than the DNA plasmids. This indicates that siRNAs are extraordinarily powerful reagents for mediating gene silencing and that siRNAs are effective at concentrations that are several orders of magnitude below the concentrations applied in conventio-

10   nal antisense or ribozyme gene targeting experiments.

In order to monitor the effect of longer dsRNAs on mammalian cells, 50 and 500 bp dsRNAs cognate to the reporter genes were prepared. As non-specific control, dsRNAs from humanized GFP (hG) (Kehlenbach, 1998)

15   was used. When dsRNAs were co-transfected, in identical amounts (not concentrations) to the siRNA duplexes, the reporter gene expression was strongly and unspecifically reduced. This effect is illustrated for HeLa cells as a representative example (Fig. 10a-d). The absolute luciferase activities were decreased unspecifically 10- to 20-fold by 50 bp dsRNA and 20- to

20   200-fold by 500 bp dsRNA co-transfection, respectively. Similar unspecific effects were observed for COS-7 and NIH/3T3 cells. For 293 cells, a 10- to 20-fold unspecific reduction was observed only for 500 bp dsRNAs. Unspecific reduction in reporter gene expression by dsRNA > 30 bp was expected as part of the interferon response.

25

Surprisingly, despite the strong unspecific decrease in reporter gene expression, we reproducibly detected additional sequence-specific, dsRNA-mediated silencing. The specific silencing effects, however, were only apparent when the relative reporter gene activities were normalized to the

30   hG dsRNA controls (Fig. 10e, f). A 2- to 10-fold specific reduction in response to cognate dsRNA was observed, also in the other three mammalian cell lines tested (data not shown). Specific silencing effects with

dsRNAs (356-1662 bp) were previously reported in CHO-K1 cells, but the amounts of dsRNA required to detect a 2- to 4-fold specific reduction were about 20-fold higher than in our experiments (Ui-Tei, 2000). Also CHO-K1 cells appear to be deficient in the interferon response. In another report,

5   293, NIH/3T3 and BHK-21 cells were tested for RNAi using luciferase/lacZ reporter combinations and 829 bp specific lacZ or 717 bp unspecific GFP dsRNA (Caplen, 2000). The failure of detecting RNAi in this case may be due to the less sensitive luciferase/lacZ reporter assay and the length differences of target and control dsRNA. Taken together, our results indi-

10   cate that RNAi is active in mammalian cells, but that the silencing effect is difficult to detect, if the interferon system is activated by dsRNA > 30 bp.

In summary, we have demonstrated for the first time siRNA-mediated gene silencing in mammalian cells. The use of short siRNAs holds great promise

15   for inactivation of gene function in human tissue culture and the develop-ment of gene-specific therapeutics.

Example 3
Specific Inhibition of Gene Expression by RNA Interference

20

3.1   Materials and Methods

3.1.1 RNA preparation and RNAi assay
Chemical RNA synthesis, annealing, and luciferase-based RNAi assays

25   were performed as described in Examples 1 or 2 or in previous publications (Tuschl et al., 1999; Zamore et al., 2000). All siRNA duplexes were direc-ted against firefly luciferase, and the luciferase mRNA sequence was derived from pGEM-luc (GenBank acc. X65316) as described (Tuschl et al., 1999). The siRNA duplexes were incubated in D. melanogaster RNAi/trans-

30   lation reaction for 15 min prior to addition of mRNAs. Translation-based RNAi assays were performed at least in triplicates.

For mapping of sense target RNA cleavage, a 177-nt transcript was generated, corresponding to the firefly luciferase sequence between positions 113-273 relative to the start codon, followed by the 17-nt complement of the SP6 promoter sequence. For mapping of antisense target RNA cleavage, a 166-nt transcript was produced from a template, which was amplified from plasmid sequence by PCR using 5' primer TAATACGACTCACTATAGAGCCCATATCGTTTCATA (T7 promoter underlined) and 3' primer AGAGGATGGAACCGCTGG. The target sequence corresponds to the complement of the firefly luciferase sequence between positions 50-215 relative to the start codon. Guanylyl transferase labelling was performed as previously described (Zamore et al., 2000). For mapping of target RNA cleavage, 100 nM siRNA duplex was incubated with 5 to 10 nM target RNA in D. melanogaster embryo lysate under standard conditions (Zamore et al., 2000) for 2 h at 25°C. The reaction was stopped by the addition of 8 volumes of proteinase K buffer (200 mM Tris-HCl pH 7.5, 25 mM EDTA, 300 mM NaCl, 2% w/v sodium dodecyl sulfate). Proteinase K (E.M. Merck, dissolved in water) was added to a final concentration of 0.6 mg/ml. The reactions were then incubated for 15 min at 65°C, extracted with phenol/chloroform/isoamyl alcohol (25:24:1) and precipitated with 3 volumes of ethanol. Samples were located on 6% sequencing gels. Length standards were generated by partial RNase T1 digestion and partial base hydrolysis of the cap-labelled sense or antisense target RNAs.

3.2   Results

3.2.1 Variation of the 3' overhang in duplexes of 21-nt siRNAs

As described above, 2 or 3 unpaired nucleotides at the 3' end of siRNA duplexes were more efficient in target RNA degradation than the respective blunt-ended duplexes. To perform a more comprehensive analysis of the function of the terminal nucleotides, we synthesized five 21-nt sense siRNAs, each displayed by one nucleotide relative to the target RNA, and eight 21-nt antisense siRNAs, each displaced by one nucleotide relative to

the target (Figure 11A). By combining sense and antisense siRNAs, eight series of siRNA duplexes with synthetic overhanging ends were generated covering a range of 7-nt 3' overhang to 4-nt 5' overhang. The interference of siRNA duplexes was measured using the dual luciferase assay system (Tuschl et al., 1999; Zamore et al., 2000). siRNA duplexes were directed against firefly luciferase mRNA, and sea pansy luciferase mRNA was used as internal control. The luminescence ratio of target to control luciferase activity was determined in the presence of siRNA duplex and was normalized to the ratio observed in the absence of dsRNA. For comparison, the interference ratios of long dsRNAs (39 to 504 pb) are shown in Figure 11B. The interference ratios were determined at concentrations of 5 nM for long dsRNAs (Figure 11A) and at 100 nM for siRNA duplexes (Figure 11C-J). The 100 nM concentrations of siRNAs was chosen, because complete processing of 5 nM 504 bp dsRNA would result in 120 nM total siRNA duplexes.

The ability of 21-nt siRNA duplexes to mediate RNAi is dependent on the number of overhanging nucleotides or base pairs formed. Duplexes with four to six 3' overhanging nucleotides were unable to mediate RNAi (Figure 11C-F), as were duplexes with two or more 5' overhanging nucleotides (Figure 11G-J). The duplexes with 2-nt 3' overhangs were most efficient in mediating RNA interference, though the efficiency of silencing was also sequence-dependent, and up to 12-fold differences were observed for different siRNA duplexes with 2-nt 3' overhangs (compare Figure 11D-H). Duplexes with blunted ends, 1-nt 5' overhang or 1- to 3-nt 3' overhangs were sometimes functional. The small silencing effect observed for the siRNA duplex with 7-nt 3' overhang (Figure 11C) may be due to an antisense effect of the long 3' overhang rather than due to RNAi. Comparison of the efficiency of RNAi between long dsRNAs (Fig. 11B) and the most effective 21-nt siRNA duplexes (Fig. 11E, G, H) indicates that a single siRNA duplex at 100 nM concentration can be as effective as 5 nM 504 bp dsRNA.

3.2.2 Length variation of the sense siRNA paired to an invariant 21-nt antisense siRNA

In order to investigate the effect of length of siRNA on RNAi, we prepared 3 series of siRNA duplexes, combining three 21-nt antisense strands with eight, 18- to 25-nt sense strands. The 3' overhang of the antisense siRNA was fixed to 1, 2, or 3 nt in each siRNA duplex series, while the sense siRNA was varied at its 3' end (Figure 12A). Independent of the lenght of the sense siRNA, we found that duplexes with 2-nt 3' overhang of anti-sense siRNA (Figure 12C) were more active than those with 1- or 3-nt 3' overhang (Figure 12B, D). In the first series, with 1-nt 3' overhang of antisense siRNA, duplexes with a 21- and 22-nt sense siRNAs, carrying a 1- and 2-nt 3' overhang of sense siRNA, respectively, were most active. Duplexes with 19- to 25-nt sense siRNAs were also able to mediate RNA, but to a lesser extent. Similarly, in the second series, with 2-nt overhang of antisense siRNA, the 21-nt siRNA duplex with 2-nt 3' overhang was most active, and any other combination with the 18- to 25-nt sense siRNAs was active to a significant degree. In the last series, with 3-nt anti-sense siRNA 3' overhang, only the duplex with a 20-nt sense siRNA and the 2-nt sense 3' overhang was able to reduce target RNA expression. Together, these results indicate that the length of the siRNA as well as the length of the 3' overhang are important, and that duplexes of 21-nt siRNAs with 2-nt 3' overhang are optimal for RNAi.

3.2.3 Length variation of siRNA duplexes with a constant 2-nt 3' overhang

We then examined the effect of simultaneously changing the length of both siRNA strands by maintaining symmetric 2-nt 3' overhangs (Figure 13A). Two series of siRNA duplexes were prepared including the 21-nt siRNA duplex of Figure 11H as reference. The length of the duplexes was varied between 20 to 25 bp by extending the base-paired segment at the 3' end of the sense siRNA (Figure 13B) or at the 3' end of the antisense siRNA (Figure 13C). Duplexes of 20 to 23 bp caused specific repression of target luciferase activity, but the 21-nt siRNA duplex was at least 8-fold more

efficient than any of the other duplexes. 24- and 25-nt siRNA duplexes did not result in any detectable interference. Sequence-specific effects were minor as variations on both ends of the duplex produced similar effects.

3.2.4 2'-Deoxy and 2'-O-methyl-modified siRNA duplexes

To assess the importance of the siRNA ribose residues for RNAi, duplexes with 21-nt siRNAs and 2-nt 3' overhangs with 2'-deoxy- or 2'-O-methyl-modified strands were examined (Figure 14). Substitution of the 2-nt 3' overhangs by 2'-deoxy nucleotides had no effect, and even the replacement of two additional riboncleotides adjacent to the overhangs in the paired region, produced significantly active siRNAs. Thus, 8 out of 42 nt of a siRNA duplex were replaced by DNA residues without loss of activity. Complete substitution of one or both siRNA strands by 2'-deoxy residues, however, abolished RNAi, as did substitution by 2'-O-methyl residues.

3.2.5 Definition of target RNA cleavage sites

Target RNA cleavage positions were previously determined for 22-nt siRNA duplexes and for a 21-nt/22-nt duplex. It was found that the position of the target RNA cleavage was located in the centre of the region covered by the siRNA duplex, 11 or 12 nt downstream of the first nucleotide that was complementary to the 21- or 22-nt siRNA guide sequence. Five distinct 21-nt siRNA duplexes with 2-nt 3' overhang (Figure 15A) were incubated with 5' cap-labelled sense or antisense target RNA in D. melanogaster lysate (Tuschl et al., 1999; Zamore et al., 2000). The 5' cleavage products were resolved on sequencing gels (Figure 15B). The amount of sense target RNA cleaved correlates with the efficiency of siRNA duplexes determined in the translation-based assay, and siRNA duplexes 1, 2 and 4 (Figure 15B and 11H, G, E) cleave target RNA faster than duplexes 3 and 5 (Figure 15B and 11F, D). Notably, the sum of radioactivity of the 5' cleavage product and the input target RNA were not constant over time, and the 5' cleavage products did not accumulate. Presumably, the cleavage products, once

- 47 -

released from the siRNA-endonuclease complex, are rapidly degraded due to the lack of either of the poly(A) tail of the 5'-cap.

The cleavage sites for both, sense and antisense target RNAs were located in the middle of the region spanned by the siRNA duplexes. The cleavage sites for each target produced by the 5 different duplexes varied by 1-nt according to the 1-nt displacement of the duplexes along the target sequences. The targets were cleaved precisely 11 nt downstream of the target position complementary to the 3'-most nucleotide of the sequence-complementary guide siRNA (Figure 15A, B).

In order to determine, whether the 5' or the 3' end of the guide siRNA sets the ruler for target RNA cleavage, we devised the experimental strategy outlined in Figure 16A and B. A 21-nt antisense siRNA, which was kept invariant for this study, was paired with sense siRNAs that were modified at either of their 5' or 3' ends. The position of sense and antisense target RNA cleavage was determined as described above. Changes in the 3' end of the sense siRNA, monitored for 1-nt 5' overhang to 6-nt 3' overhang, did neither effect the position of sense nor antisense target RNA cleavage (Figure 16C). Changes in the 5' end of the sense siRNA did no affect the sense target RNA cleavage (Figure 16D, top panel), which was expected because the antisense siRNA was unchanged. However, the antisense target RNA cleavage was affected and strongly dependent on the 5' end of the sense siRNA (Figure 16D, bottom panel). The antisense target was only cleaved, when the sense siRNA was 20 or 21 nt in size, and the position of cleavage different by 1-nt, suggesting that the 5' end of the target-recognizing siRNA sets the ruler for target RNA cleavage. The position is located between nucleotide 10 and 11 when counting in upstream direction from the target nucleotide paired to the 5'-most nucleotide of the guide siRNA (see also Figure 15A).

3.2.6 Sequence effects and 2'-deoxy substitutions in the 3' overhang

A 2-nt 3'overhang is preferred for siRNA function. We wanted to know, if the sequence of the overhanging nucleotides contributes to target recognition, or if it is only a feature required for reconstitution of the endonuclease complex (RISC or siRNP). We synthesized sense and antisense siRNAs with AA, CC, GG, UU, and UG 3' overhangs and included the 2'-deoxy modifications TdG and TT. The wild-type siRNAs contained AA in the sense 3' overhang and UG in the antisense 3' overhang (AA/UG). All siRNA duplexes were functional in the interference assay and reduced target expression at least 5-fold (Figure 17). The most efficient siRNA duplexes that reduced target expression more than 10-fold, were of the sequence type NN/UG, NN/UU, NN/TdG, and NN/TT (N, any nucleotide). siRNA duplexes with an antisense siRNA 3' overhang of AA, CC or GG were less active by a factor 2 to 4 when compared to the wild-type sequence UG or the mutant UU. This reduction in RNAi efficiency is likely due to the contribution of the penultimate 3' nucleotide to sequence-specific target recognition, as the 3' terminal nucleotide was changed from G to U without effect.

Changes in the sequence of the 3' overhang of the sense siRNA did not reveal any sequence-dependent effects, which was expected, because the sense siRNA must not contribute to sense target mRNA recognition.

3.2.7 Sequence specifity of target recognition

In order to examine the sequence-specifity of target recognition, we introduced sequence changes into the paired segments of siRNA duplexes and determined the efficiency of silencing. Sequence changes were introduced by inverting short segments of 3- or 4-nt length or as point mutations (Figure 18). The sequence changes in one siRNA strand were compensated in the complementary siRNA strand to avoid pertubing the base-paired siRNA duplex structure. The sequence of all 2-nt 3' overhangs was TT (T, 2'-deoxythymidine) to reduce costs of synthesis. The TT/TT refe-

rence siRNA duplex was comparable in RNAi to the wild-type siRNA duplex AA/UG (Figure 17). The ability to mediate reporter mRNA destruction was quantified using the translation-based luminescence assay. Duplexes of siRNAs with inverted sequence segments showed dramatically reduced ability for targeting the firefly luciferase reporter (Figure 18). The sequence changes located between the 3' end and the middle of the antisense siRNA completely abolished target RNA recognition, but mutations near the 5' end of the antisense siRNA exhibit a small degree of silencing. Transversion of the A/U base pair located directly opposite of the predicted target RNA cleavage site, or one nucleotide further away from the predicted site, prevented target RNA cleavage, therefore indicating that single mutation within the centre of a siRNA duplex discriminate between mismatched targets.

3.3 Discussion

siRNAs are valuable reagents for inactivation of gene expression, not only in insect cells, but also in mammalian cells, with a great potential for therapeutic application. We have systematically analysed the structural determinants of siRNA duplexes required to promote efficient target RNA degradation in D. melanogaster embryo lysate, thus providing rules for the design of most potent siRNA duplexes. A perfect siRNA duplex is able to silence gene expression with an efficiency comparable to a 500 bp dsRNA, given that comparable quantities of total RNA are used.

3.4 The siRNA user guide

Efficiently silencing siRNA duplexes are preferably composed of 21-nt antisense siRNAs, and should be selected to form a 19 bp double helix with 2-nt 3' overhanging ends. 2'-deoxy substitutions of the 2-nt 3' over-hanging ribonucleotides do not affect RNAi, but help to reduce the costs of RNA synthesis and may enhance RNAse resistance of siRNA duplexes. More extensive 2'-deoxy or 2'-O-methyl modifications, however, reduce

the ability of siRNAs to mediate RNAi, probably by interfering with protein association for siRNAP assembly.

Target recognition is a highly sequence-specific process, mediated by the siRNA complementary to the target. The 3'-most nucleotide of the guide siRNA does not contribute to specificity of target recognition, while the penultimate nucleotide of the 3' overhang affects target RNA cleavage, and a mismatch reduces RNAi 2- to 4-fold. The 5' end of a guide siRNA also appears more permissive for mismatched target RNA recognition when compared to the 3' end. Nucleotides in the centre of the siRNA, located opposite the target RNA cleavage site, are important specificity determinants and even single nucleotide changes reduce RNAi to undetectable level. This suggests that siRNA duplexes may be able to discriminate mutant or polymorphic alleles in gene targeting experiments, which may become an important feature for future therapeutic developments.

Sense and antisense siRNAs, when associated with the protein components of the endonclease complex or its commitment complex, were suggested to play distinct roles; the relative orientation of the siRNA duplex in this complex defines which strand can be used for target recognition. Synthetic siRNA duplexes have dyad symmetry with respect to the double-helical structure, but not with respect to sequence. The association of siRNA duplexes with the RNAi proteins in the D. melanogaster lysate will lead to formation of two asymmetric complexes. In such hypothetical complexes, the chiral environment is distinct for sense and antisense siRNA, hence their function. The prediction obviously does not apply to palindromic siRNA sequences, or to RNAi proteins that could associate as homodimers. To minimize sequence effects, which may affect the ratio of sense and antisense-targeting siRNPs, we suggest to use siRNA sequences with identical 3' overhanging sequences. We recommend to adjust the sequence of the overhang of the sense siRNA to that of the antisense 3' overhang, because the sense siRNA does not have a target in typical

knock-down experiments. Asymmetry in reconstitution of sense and anti-sense-cleaving siRNPs could be (partially) responsible for the variation in RNAi efficiency observed for various 21-nt siRNA duplexes with 2-nt 3′ overhangs used in this study (Figure 14). Alternatively, the nucleotide sequence at the target site and/or the accessibility of the target RNA structure may be responsible for the variation in efficiency for these siRNA duplexes.

- 52 -

References

Bass, B. L. (2000). Double-stranded RNA as a template for gene silencing. Cell 101, 235-238.

Bosher, J. M., and Labouesse, M. (2000). RNA interference: genetic wand and genetic watchdog. Nat. Cell Biol. 2, E31-36.

Caplen, N. J., Fleenor, J., Fire, A., and Morgan, R. A. (2000). dsRNA-mediated gene silencing in cultured Drosophila cells: a tissue culture model for the analysis of RNA interference. Gene 252, 95-105.

Catalanotto, C., Azzalin, G., Macino, G., and Cogoni, C. (2000). Gene silencing in worms and fungi. Nature 404, 245.

Chanfreau, G., Buckle, M., and Jacquier, A. (2000). Recognition of a conserved class of RNA tetraloops by Saccharomyces cerevisiae RNase III. Proc. Natl. Acad. Sci. USA 97, 3142-3147.

Clemens, M. J. (1997). PKR--a protein kinase regulated by double-stranded RNA. Int. J. Biochem. Cell Biol. 29, 945-949.

Cogoni, C., and Macino, G. (1999). Homology-dependent gene silencing in plants and fungi: a number of variations on the same theme. Curr. Opin. Microbiol. 2, 657-662.

Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D. C. (2000). An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell 101, 543-553.

Dernburg, A. F., Zalevsky, J., Colaiacovo, M. P., and Villeneuve, A. M. (2000). Transgene-mediated cosuppression in the C. elegans germ line. Genes & Dev. 14, 1578-1583.

5 Dunn, J. J. (1982). Ribonuclease III. In The enzymes, vol 15, part B, P. D. Boyer, ed. (New York: Academic Press), pp. 485-499.

Filippov, V., Solovyev, V., Filippova, M., and Gill, S. S. (2000). A novel type of RNase III family proteins in eukaryotes. Gene 245, 213-221.

10

Fire, A. (1999). RNA-triggered gene silencing. Trends Genet. 15, 358-363.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-
15 stranded RNA in Caenorhabditis elegans. Nature 391, 806-811.

Grishok, A., Tabara, H., and Mello, C. C. (2000). Genetic requirements for inheritance of RNAi in C. elegans. Science 287, 2494-2497.

20 Hamilton, A. J., and Baulcombe, D. C. (1999). A species of small anti-sense RNA in posttranscriptional gene silencing in plants. Science 286, 950-952.

Hammond, S. M., Bernstein, E., Beach, D., and Hannon, G. J. (2000). An
25 RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature 404, 293-296.

Jacobsen, S. E., Running, M. P., and M., M. E. (1999). Disruption of an RNA helicase/RNase III gene in Arabidopsis causes unregulated cell division
30 in floral meristems. Development 126, 5231-5243.

- 54 -

Jensen, S., Gassama, M. P., and Heidmann, T. (1999). Taming of transposable elements by homology-dependent gene silencing. Nat. Genet. 21, 209-212.

Kehlenbach, R. H., Dickmanns, A. & Gerace, L. (1998). Nucleocytoplasmic shuttling factors including Ran and CRM1 mediate nuclear export of NFAT In vitro. J. Cell Biol. 141, 863-874.

Kennerdell, J. R., and Carthew, R. W. (1998). Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. Cell 95, 1017-1026.

Ketting, R. F., Haverkamp, T. H., van Luenen, H. G., and Plasterk, R. H. (1999). Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell 99, 133-141.

Ketting, R. F., and Plasterk, R. H. (2000). A genetic link between co-suppression and RNA interference in C. elegans. Nature 404, 296-298.

Lucy, A. P., Guo, H. S., Li, W. X., and Ding, S. W. (2000). Suppression of post-transcriptional gene silencing by a plant viral protein localized in the nucleus. EMBO J. 19, 1672-1680.

Matsuda, S., Ichigotani, Y., Okuda, T., Irimura, T., Nakatsugawa, S., and Hamaguchi, M. (2000). Molecular cloning and characterization of a novel human gene (HERNA) which encodes a putative RNA-helicase. Biochim. Biophys. Acta 31, 1-2.

Milligan, J.F., and Uhlenbeck, O.C. (1989). Synthesis of small RNAs using T7 RNA polymerase. Methods Enzymol. 180, 51-62.

Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J. B., Jouette, D., Lacombe, A. M., Nikic, S., Picault, N., Remoue, K., Sanial, M., Vo, T. A., and Vaucheret, H. (2000). Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. Cell 101, 533-542.

Ngo, H., Tschudi, C., Gull, K., and Ullu, E. (1998). Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. Proc. Natl. Acad. Sci. USA 95, 14687-14692.

Nicholson, A. W. (1999). Function, mechanism and regulation of bacterial ribonucleases. FEMS Microbiol. Rev. 23, 371-390.

Oelgeschlager, M., Larrain, J., Geissert, D., and De Robertis, E. M. (2000). The evolutionarily conserved BMP-binding protein Twisted gastrulation promotes BMP signalling. Nature 405, 757-763.

Pan, T., and Uhlenbeck, O. C. (1992). In vitro selection of RNAs that undergo autolytic cleavage with Pb^{2+}. Biochemistry 31, 3887-3895.

Pelissier, T., and Wassenegger, M. (2000). A DNA target of 30 bp is sufficient for RNA-directed methylation. RNA 6, 55-65.

Plasterk, R. H., and Ketting, R. F. (2000). The silence of the genes. Curr. Opin. Genet. Dev. 10, 562-567.

Ratcliff, F. G., MacFarlane, S. A., and Baulcombe, D. C. (1999). Gene Silencing without DNA. RNA-mediated cross-protection between viruses. Plant Cell 11, 1207-1216.

Robertson, H. D. (1990). Escherichia coli ribonuclease III. Methods Enzymol. 181, 189-202.

Robertson, H. D. (1982). Escherichia coli ribonuclease III cleavage sites. Cell 30, 669-672.

Romaniuk, E., McLaughlin, L. W., Neilson, T., and Romaniuk, P. J. (1982). The effect of acceptor oligoribonucleotide sequence on the T4 RNA ligase reaction. Eur J Biochem 125, 639-643.

Sharp, P. A. (1999). RNAi and double-strand RNA. Genes & Dev. 13, 139-141.

Sijen, T., and Kooter, J. M. (2000). Post-transcriptional gene-silencing: RNAs on the attack or on the defense? Bioessays 22, 520-531.

Smardon, A., Spoerke, J., Stacey, S., Klein, M., Mackin, N., and Maine, E. (2000). EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. Curr. Biol. 10, 169-178.

Svoboda, P., Stein, P., Hayashi, H., and Schultz, R. M. (2000). Selective reduction of dormant maternal mRNAs in mouse oocytes by RNA interference. Development 127, 4147-4156.

Tabara, H., Sarkissian, M., Kelly, W. G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C. C. (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell 99, 123-132.

Tuschl, T., Ng, M. M., Pieken, W., Benseler, F., and Eckstein, F. (1993). Importance of exocyclic base functional groups of central core guanosines for hammerhead ribozyme activity. Biochemistry 32, 11658-11668.

- 57 -

Tuschl, T., Sharp, P. A., and Bartel, D. P. (1998). Selection in vitro of novel ribozymes from a partially randomized U2 and U6 snRNA library. EMBO J. 17, 2637-2650.

Tuschl, T., Zamore, P. D., Lehmann, R., Bartel, D. P., and Sharp, P. A. (1999). Targeted mRNA degradation by double-stranded RNA in vitro. Genes & Dev. 13, 3191-3197.

Ui-Tei, K., Zenno, S., Miyata, Y. & Saigo, K. (2000). Sensitive assay of RNA interference in Drosophila and Chinese hamster cultured cells using firefly luciferase gene as target. FEBS Letters 479, 79-82.

Verma, S., and Eckstein, F. (1999). Modified oligonucleotides: Synthesis and strategy for users. Annu. Rev. Biochem. 67, 99-134.

Voinnet, O., Lederer, C., and Baulcombe, D. C. (2000). A viral movement protein prevents spread of the gene silencing signal in Nicotiana ben-thamiana. Cell 103, 157-167.

Wassenegger, M. (2000). RNA-directed DNA methylation. Plant Mol. Biol. 43, 203-220.

Wianny, F., and Zernicka-Goetz, M. (2000). Specific interference with gene function by double-stranded RNA in early mouse development. Nat. Cell Biol. 2, 70-75.

Wu, H., Xu, H., Miraglia, L. J., and Crooke, S. T. (2000). Human RNase III is a 160 kDa Protein Involved in Preribosomal RNA Processing. J. Biol. Chem. 17, 17.

- 58 -

Yang, D., Lu, H. and Erickson, J.W. (2000) Evidence that processed small dsRNAs may mediate sequence-specific mRNA degradation during RNAi in drosophilia embryos. Curr. Biol., 10, 1191-1200.

Zamore, P. D., Tuschl, T., Sharp, P. A., and Bartel, D. P. (2000). RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell 101, 25-33.

Zhang, K., and Nicholson, A. W. (1997). Regulation of ribonuclease III processing by double-helical sequence antideterminants. Proc. Natl. Acad. Sci. USA 94, 13437-13441.

Claims

1. Isolated double-stranded RNA molecule, wherein each RNA strand has a length from 19-25 nucleotides, wherein said RNA molecule is capable of target-specific nucleic acid modifications.

2. The RNA molecule of claim 1 wherein at least one strand has a 3'-overhang from 1-5 nucleotides.

3. The RNA molecule of claim 1 or 2 capable of target-specific RNA interference and/or DNA methylation.

4. The RNA molecule of any one of claims 1-3, wherein each strand has a length from 19-23, particularly from 20-22 nucleotides.

5. The RNA molecule of any one of claims 2-4, wherein the 3'-overhang is from 1-3 nucleotides.

6. The RNA molecule of any one of claims 2-5, wherein the 3'-overhang is stabilized against degradation.

7. The RNA molecule of any one of claims 1-6, which contains at least one modified nucleotide analogue.

8. The RNA molecule of claim 7, wherein the modified nucleotide analogue is selected from sugar- or backbone modified ribonucleotides.

9. The RNA molecule according to claim 7 or 8, wherein the nucleotide analogue is a sugar-modified ribonucleotide, wherein the 2'-OH group is replaced by a group selected from H, OR, R, halo, SH, SR^1,

- 60 -

NH$_2$, NHR, NR$_2$ or CN, wherein R is C$_1$-C$_6$ alkyl, alkenyl or alkynyl and halo is F, Cl, Br or I.

10. The RNA molecule of claim 7 or 8, wherein the nucleotide analogue is a backbone-modified ribonucleotide containing a phosphothioate group.

11. The RNA molecule of any one of claims 1-10, which has a sequence having an identity of at least 50 percent to a predetermined mRNA target molecule.

12. The RNA molecule of claim 11, wherein the identity is at least 70 percent.

13. A method of preparing a double-stranded RNA molecule of any one of claims 1-12 comprising the steps:

(a) synthesizing two RNA strands each having a length from 19-25 nucleotides, wherein said RNA strands are capable of forming a double-stranded RNA molecule,

(b) combining the synthesized RNA strands under conditions, wherein a double-stranded RNA molecule is formed, which is capable of target-specific nucleic acid modifications.

14. The method of claim 13, wherein the RNA strands are chemically synthesized.

15. The method of claim 13, wherein the RNA strands are enzymatically synthesized.

16.   A method of mediating target-specific nucleic acid modifications in a cell or an organism comprising the steps:

    (a)   contacting said cell or organism with the double-stranded RNA molecule of any one of claims 1-12 under conditions wherein target-specific nucleic acid modifications can occur, and

    (b)   mediating a target-specific nucleic acid modification effected by the double-stranded RNA towards a target nucleic acid having a sequence portion substantially corresponding to the double-stranded RNA.

17.   The method of claim 16, wherein the nucleic acid modification is RNA interference and/or DNA methylation.

18.   The method of claim 16 and 17 wherein said contacting comprises introducing said double-stranded RNA molecule into a target cell in which the target-specific nucleic acid modification can occur.

19.   The method of claim 18 wherein the introducing comprises a carrier-mediated delivery or injection.

20.   Use of the method of any one of claims 16-19 for determining the function of a gene in a cell or an organism.

21.   Use of the method of any one of claims 16-19 for modulating the function of a gene in a cell or an organism.

22.   The use of claim 20 or 21, wherein the gene is associated with a pathological condition.

23.   The use of claim 22, wherein the gene is a pathogen-associated gene.

24.   The use of claim 23, wherein the gene is a viral gene.

25.   The use of claim 22, wherein the gene is a tumor-associated gene.

26.   The use of claim 22, wherein the gene is an autoimmune disease-associated gene.

27.   Pharmaceutical composition containing as an active agent at least one double-stranded RNA molecule of any one of claims 1-12 and a pharmaceutical carrier.

28.   The composition of claim 27 for diagnostic applications.

29.   The composition of claim 27 for therapeutic applications.

30.   A eukaryotic cell or a eukaryotic non-human organism exhibiting a target gene-specific knockout phenotype wherein said cell or organism is transfected with at least one double-stranded RNA molecule capable of inhibiting the expression of an endogeneous target gene or with a DNA encoding at least one double-stranded RNA molecule capable of inhibiting the expression of at least one endogeneous target gene.

31.   The cell or organism of claim 30 which is a mammalian cell.

32.   The cell or organism of claim 31 which is a human cell.

33.   The cell or organism of any one of claims 30-32 which is further transfected with at least one exogeneous target nucleic acid coding

for the target protein or a variant or mutated form of the target protein, wherein said exogeneous target nucleic acid differs from the endogeneous target gene on the nucleic acid level such that the expression of the exogeneous target nucleic acid is substantially less inhibited by the double stranded RNA molecule than the expression of the endogeneous target gene.

34.   The cell or organism of claim 33 wherein the exogeneous target nucleic acid is fused to a further nucleic acid sequence encoding a detectable peptide or polypeptide.

35.   Use of the cell or organism of any of claims 30-34 for analytic procedures.

36.   The use of claim 35 for the analysis of gene expression profiles.

37.   The use of claim 35 for a proteome analysis.

38.   The use of any one of claims 35-37 wherein an analysis of a variant or mutant form of the target protein encoded by an exogeneous target nucleic acid is carried out.

39.   The use of claim 38 for identifying functional domains of the target protein.

40.   The use of any one of claims 35-39 wherein a comparison of at least two cells or organisms is carried out selected from:
      (i)   a control cell or control organism without target gene inhibition,
      (ii)   a cell or organism with target gene inhibition and
      (iii) a cell or organism with target gene inhibition plus target gene complementation by an exogeneous target nucleic acid.

41.   The use of any one of claims 35-40 wherein the analysis comprises a functional and/or phenotypic analysis.

42.   Use of a cell of any one of claims 30-34 for preparative procedures.

43.   The use of claim 41 for the isolation of proteins or protein complexes from eukaryotic cells.

44.   The use of claim 43 for the isolation of high molecular weight protein complexes which may optionally contain nucleic acids.

45.   The use of any one of claims 35-44 in a procedure for identifying and/or characterizing pharmacological agents.

46.   A system for identifying and/or characterizing a pharmacological agent acting on at least one target protein comprising:
       (a)   a eukaryotic cell or a eukaryotic non-human organism capable of expressing at least one target gene coding for said at least one target protein,
       (b)   at least one double-stranded RNA molecule capable of inhibiting the expression of said at least one endogeneous target gene, and
       (c)   a test substance or a collection of test substances wherein pharmacological properties of said test substance or said collection are to be identified and/or characterized.

47.   The system of claim 46 further comprising:
       (d)   at least one exogeneous target nucleic acid coding for the target protein or a variant or mutated from of the target protein wherein said exogeneous target nucleic acid differs from the endogeneous target gene on the nucleic acid level such that the expression of the exogeneous target nucleic

acid is substantially less inhibited by the double stranded RNA molecule than the expression of the endogeneous target gene.

FIGURE 1A
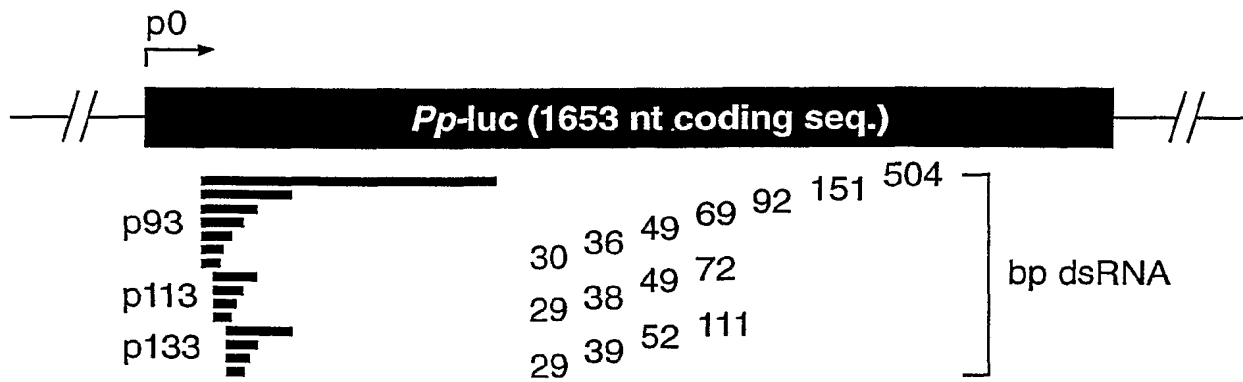
FIGURE 1B
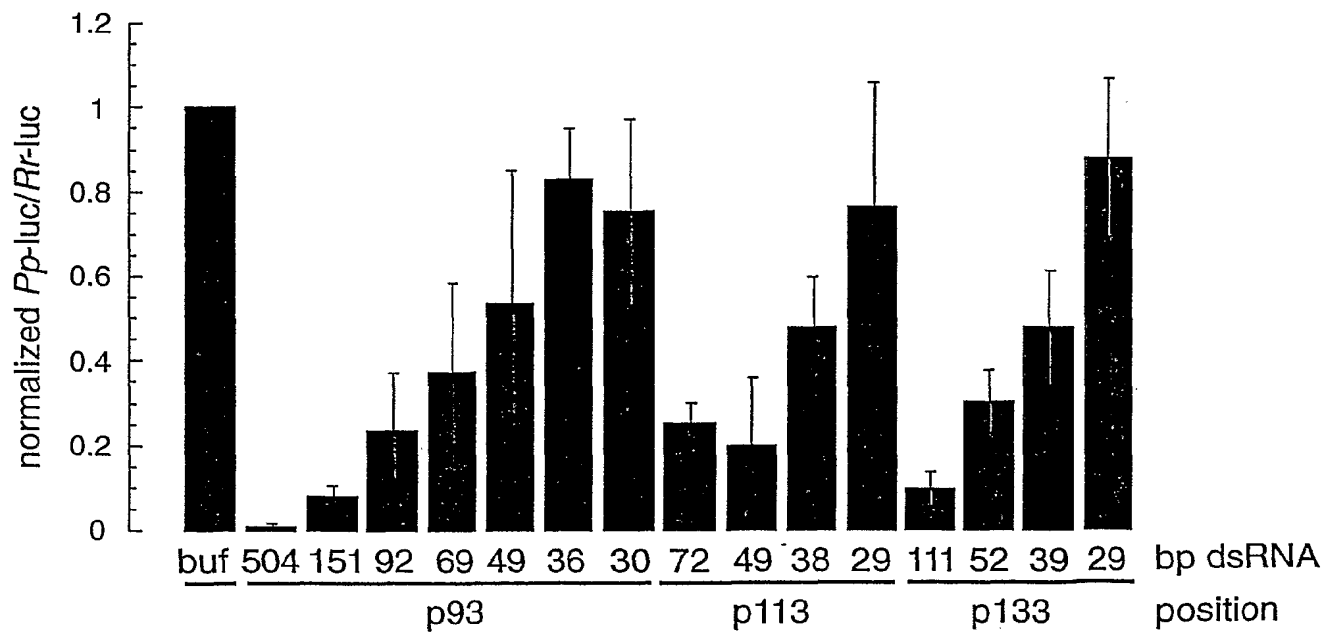
FIGURE 2

Figure 3A

FIGURE 3B

|←————10 nt————→|

sense target

5′ 7mG*pppGAACAAUUGCUUUUACAGAUGCACAUAUCGAGGUG

antisense target   **dsRNAs**

3′ CGCAUUAUGCUGAGUGAUAUCUUGUUAACGAAAAUGUCUACGUGUAUAGCUCCAC

AACAUCACGUACGCGGAAUACUUCGAAAUGUCCGUUCGGUUGGCAGAAGCUAUGAAAC

39 ———— 52

UUGUAGUGCAUGCGCCUUAUGAAGCUUUACAGGCAAGCCAACCGUCUUCGAUACUUUG

|←——10 nt——→| |←——10 nt——→|

GAUAUGGGCUGAAUACAAAUCACAGAAUCGUCGUAUGCAGUGAAAACUCUCUUCAAUU

————111

CUAUACCCGACUUAUGUUUAGUGUCUUAGCAGCAUACGUCACUUUUGAGAGAAGUUAA

|←——9 nt——→|

CUUUAUGCCUAUAGUGUCACCUAAAU 3′

GAAAUACGGppp*Gm7 5′

FIGURE 4A

39 bp ds*Pp*-luc p133

5 ' GCACAUAUCGAGGUGAACAUCACGUACGCGGAAUACUUC
3 ' CGUGUAUAGCUCCACUUGUAGUGCAUGCGCCUUAUGAAG

52 bp ds*Pp*-luc p133

5 ' GCACAUAUCGAGGUGAACAUCACGUACGCGGAAUACUUCGAAAUGUCCGUUC
3 ' CGUGUAUAGCUCCACUUGUAGUGCAUGCGCCUUAUGAAGCUUUACAGGCAAG

111 bp ds*Pp*-luc p133

5 ' GCACAUAUCGAGGUGAACAUCACGUACGCGGAAUACUUCGAAAUGUCCGUUCGGU
3 ' CGUGUAUAGCUCCACUUGUAGUGCAUGCGCCUUAUGAAGCUUUACAGGCAAGCCA

UGGCAGAAGCUAUGAAACGAUAUGGGCUGAAUACAAAUCACAGAAUCGUCGUAUGC
ACCGUCUUCGAUACUUUGCUAUACCCGACUUAUGUUUAGUGUCUUAGCAGCAUACG

FIGURE 4B

Figure 5A

siRNAs/
dsRNA

ds52   5' GCACAUAUCGAGGUGAACAUCACGUACGCGGAAUACUUCGAAAUGUCCGUUC
(control) 3' CACGUGUAUAGCUCCACUUGUAGUGCAUGCGCCUUAUGAAGCUUUACAGGCAAG

1

2

3

4

5

6

7

FIGURE 5B

FIGURE 6A

FIGURE 6B

Figure 7

a

SV40 prom.   3'-UTR   enh.

Pp-luc (GL2-SV40)

SV40 prom.   poly(A)   enh.

Pp-luc (GL3-SV40)

TK prom.   poly(A)

Rr-luc (RL-TK)
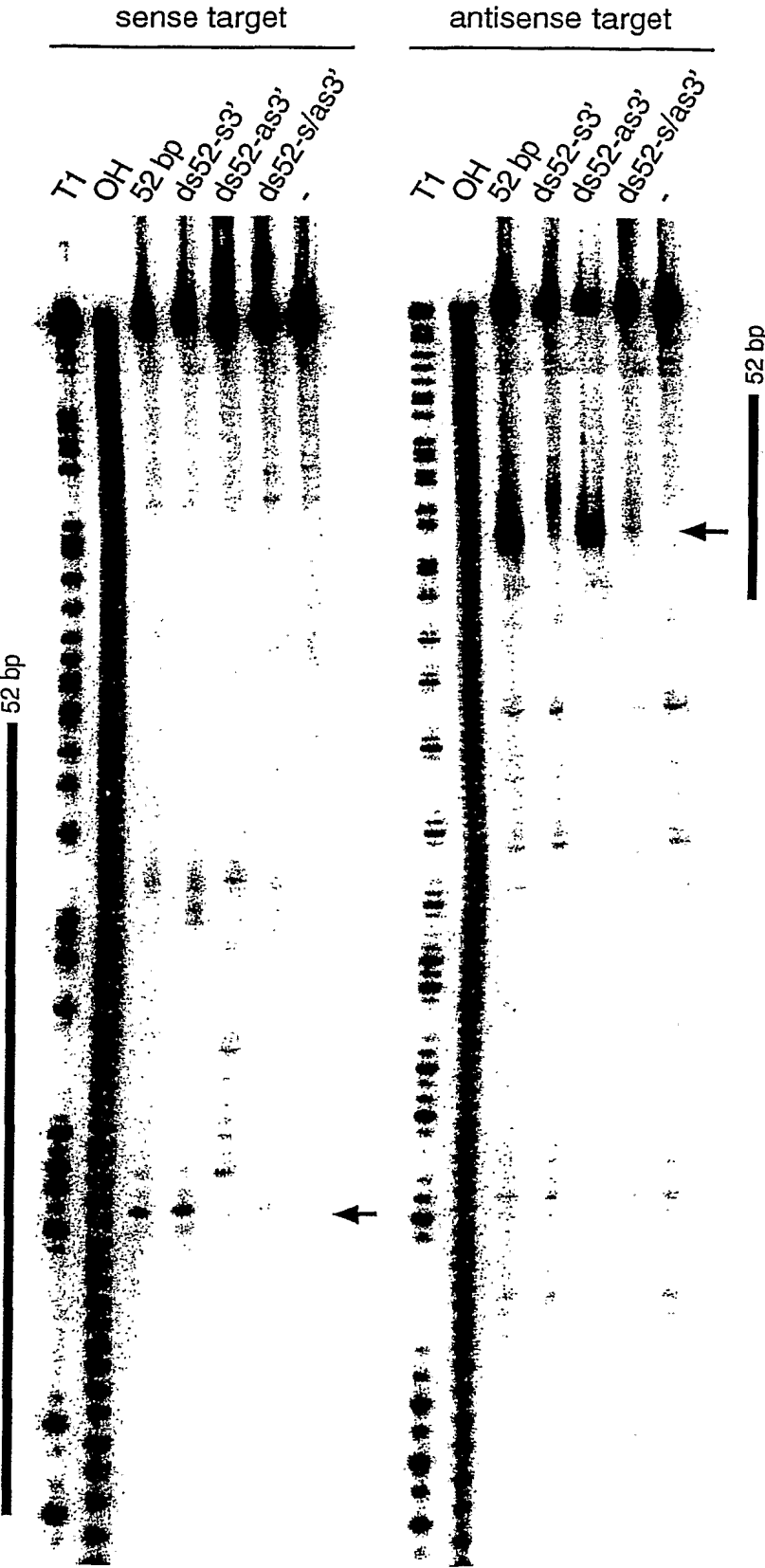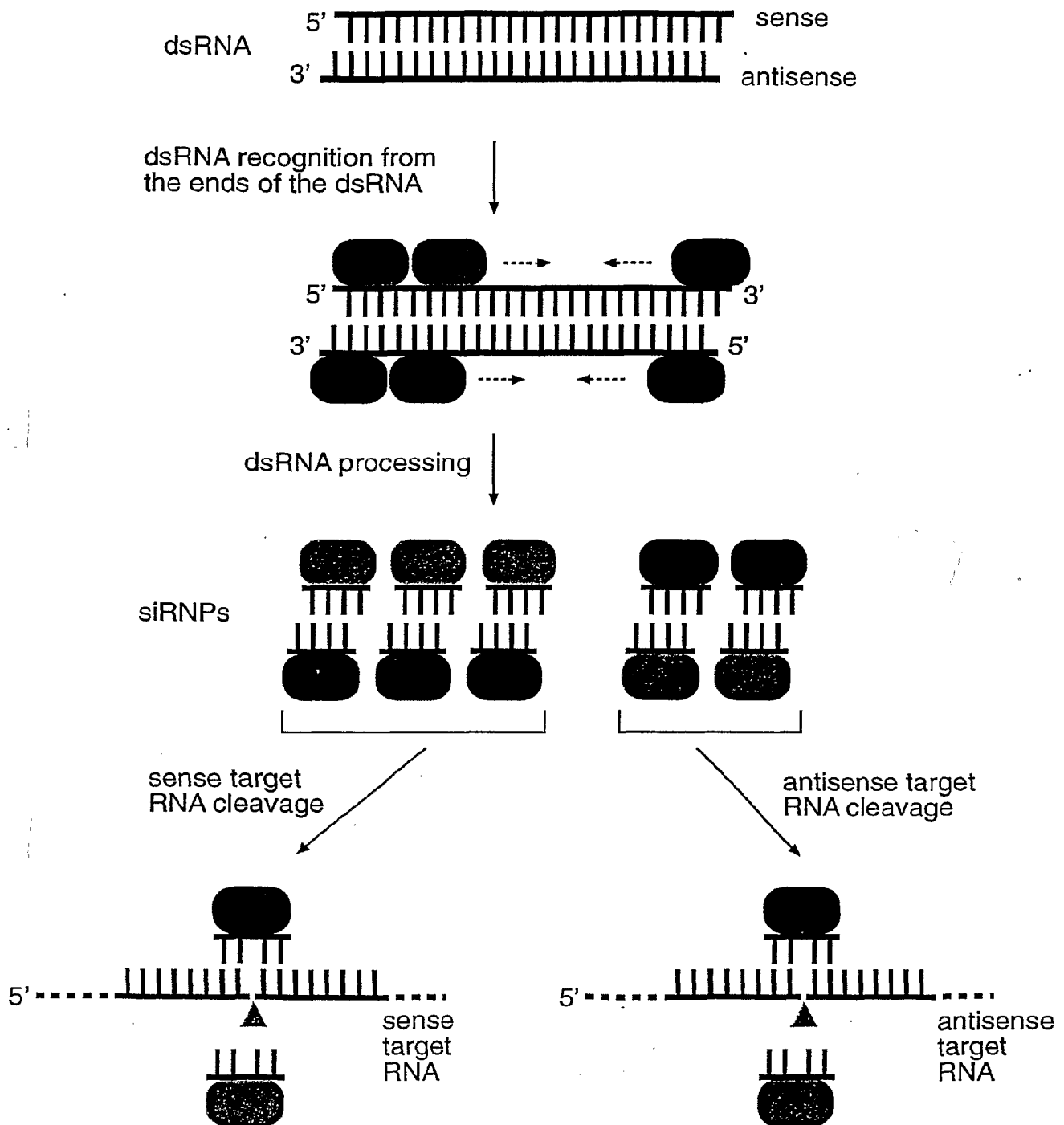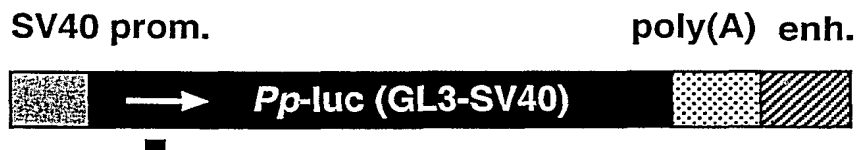
b

siRNA
duplex

uGL2   5' CGUACGCGGAAUACUUCGAUU
            UUGCAUGCGCCUUAUGAAGCU 5'

GL2   5' CGUACGCGGAAUACUUCGATT
            TTGCAUGCGCCUUAUGAAGCU 5'

GL3   5' CUUACGCUGAGUACUUCGATT
            TTGAAUGCGACUCAUGAAGCU 5'

invGL2   5' AGCUUCAUAAGGCGCAUGCTT
            TTUCGAAGUAUUCCGCGUACG 5'

RL   5' AAACAUGCAGAAAAUGCUGTT
            TTUUUGUACGUCUUUUACGAC 5'

Figure 8

Figure 9

Figure 10

Figure 11

A

sense siRNAs   sense target

7mG ——————————————AAAAAAAA

antisense siRNAs

3 nt   3' overhang
4 nt
5 nt
6 nt
7 nt

-4 nt
-3 nt
-2 nt
-1 nt
0 nt

B

5 nM

norm. Pp-luc/Rr-luc

bu 504 111 52 39
base pairs

C

5' AUCACGUACGCGGAAUACUUC
   UUGUAGUGCAUGCGCCUUAUG 5'

5' UCACGUACGCGGAAUACUUCG
   UUGUAGUGCAUGCGCCUUAUG 5'

5' CACGUACGCGGAAUACUUCGA
   UUGUAGUGCAUGCGCCUUAUG 5'

5' ACGUACGCGGAAUACUUCGAA
   UUGUAGUGCAUGCGCCUUAUG 5'

5' CGUACGCGGAAUACUUCGAAA
   UUGUAGUGCAUGCGCCUUAUG 5'

100 nM

3' overhang (nt)

D

5' AUCACGUACGCGGAAUACUUC
   UGUAGUGCAUGCGCCUUAUGA 5'

5' UCACGUACGCGGAAUACUUCG
   UGUAGUGCAUGCGCCUUAUGA 5'

5' CACGUACGCGGAAUACUUCGA
   UGUAGUGCAUGCGCCUUAUGA 5'

5' ACGUACGCGGAAUACUUCGAA
   UGUAGUGCAUGCGCCUUAUGA 5'

5' CGUACGCGGAAUACUUCGAAA
   UGUAGUGCAUGCGCCUUAUGA 5'

2 3 4 5 6

E

5' AUCACGUACGCGGAAUACUUC
   GUAGUGCAUGCGCCUUAUGAA 5'

5' UCACGUACGCGGAAUACUUCG
   GUAGUGCAUGCGCCUUAUGAA 5'

5' CACGUACGCGGAAUACUUCGA
   GUAGUGCAUGCGCCUUAUGAA 5'

5' ACGUACGCGGAAUACUUCGAA
   GUAGUGCAUGCGCCUUAUGAA 5'

5' CGUACGCGGAAUACUUCGAAA
   GUAGUGCAUGCGCCUUAUGAA 5'

1 2 3 4 5

F

5' AUCACGUACGCGGAAUACUUC
   UAGUGCAUGCGCCUUAUGAAG 5'

5' UCACGUACGCGGAAUACUUCG
   UAGUGCAUGCGCCUUAUGAAG 5'

5' CACGUACGCGGAAUACUUCGA
   UAGUGCAUGCGCCUUAUGAAG 5'

5' ACGUACGCGGAAUACUUCGAA
   UAGUGCAUGCGCCUUAUGAAG 5'

5' CGUACGCGGAAUACUUCGAAA
   UAGUGCAUGCGCCUUAUGAAG 5'

0 1 2 3 4

G

5' AUCACGUACGCGGAAUACUUC
   AGUGCAUGCGCCUUAUGAAGC 5'

5' UCACGUACGCGGAAUACUUCG
   AGUGCAUGCGCCUUAUGAAGC 5'

5' CACGUACGCGGAAUACUUCGA
   AGUGCAUGCGCCUUAUGAAGC 5'

5' ACGUACGCGGAAUACUUCGAA
   AGUGCAUGCGCCUUAUGAAGC 5'

5' CGUACGCGGAAUACUUCGAAA
   AGUGCAUGCGCCUUAUGAAGC 5'

-1 0 1 2 3

H

5' AUCACGUACGCGGAAUACUUC
   GUGCAUGCGCCUUAUGAAGCU 5'

5' UCACGUACGCGGAAUACUUCG
   GUGCAUGCGCCUUAUGAAGCU 5'

5' CACGUACGCGGAAUACUUCGA
   GUGCAUGCGCCUUAUGAAGCU 5'

5' ACGUACGCGGAAUACUUCGAA
   GUGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGAAA
   GUGCAUGCGCCUUAUGAAGCU 5'

-2 -1 0 1 2

I

5' AUCACGUACGCGGAAUACUUC
   UGCAUGCGCCUUAUGAAGCUU 5'

5' UCACGUACGCGGAAUACUUCG
   UGCAUGCGCCUUAUGAAGCUU 5'

5' CACGUACGCGGAAUACUUCGA
   UGCAUGCGCCUUAUGAAGCUU 5'

5' ACGUACGCGGAAUACUUCGAA
   UGCAUGCGCCUUAUGAAGCUU 5'

5' CGUACGCGGAAUACUUCGAAA
   UGCAUGCGCCUUAUGAAGCUU 5'

-3 -2 -1 0 1

J

5' AUCACGUACGCGGAAUACUUC
   GCAUGCGCCUUAUGAAGCUUU 5'

5' UCACGUACGCGGAAUACUUCG
   GCAUGCGCCUUAUGAAGCUUU 5'

5' CACGUACGCGGAAUACUUCGA
   GCAUGCGCCUUAUGAAGCUUU 5'

5' ACGUACGCGGAAUACUUCGAA
   GCAUGCGCCUUAUGAAGCUUU 5'

5' CGUACGCGGAAUACUUCGAAA
   GCAUGCGCCUUAUGAAGCUUU 5'

-4 -3 -2 -1 0

A

```
                                        -2 to 7 nt
                                        3' overhang
                                          ┊        ┊
sense siRNA (18-25 nt)        ▬▬▬▬▬▬▬▬▬▬ ▬ ▬ ▬ ▬ ──→
antisense siRNA (21 nt)     ▪▪▬▬▬▬▬▬▬▬▬▬▪▪▪▪▪▪▪ ──→
                            ┊ ┊
                            1 to 3 nt
                            3' overhang
```

B

```
5' CGUACGCGGAAUACUUCG
   UGCAUGCGCCUUAUGAAGCUU 5'

5' CGUACGCGGAAUACUUCGA
   UGCAUGCGCCUUAUGAAGCUU 5'

5' CGUACGCGGAAUACUUCGAA
   UGCAUGCGCCUUAUGAAGCUU 5'

5' CGUACGCGGAAUACUUCGAAA
   UGCAUGCGCCUUAUGAAGCUU 5'

5' CGUACGCGGAAUACUUCGAAAU
   UGCAUGCGCCUUAUGAAGCUU 5'

5' CGUACGCGGAAUACUUCGAAAUG
   UGCAUGCGCCUUAUGAAGCUU 5'

5' CGUACGCGGAAUACUUCGAAAUGU
   UGCAUGCGCCUUAUGAAGCUU 5'

5' CGUACGCGGAAUACUUCGAAAUGUC
   UGCAUGCGCCUUAUGAAGCUU 5'
```
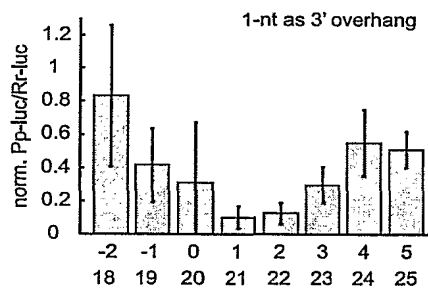
C

```
5' CGUACGCGGAAUACUUCG
   GUGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGA
   GUGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGAA
   GUGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGAAA
   GUGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGAAAU
   GUGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGAAAUG
   GUGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGAAAUGU
   GUGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGAAAUGUC
   GUGCAUGCGCCUUAUGAAGCU 5'
```
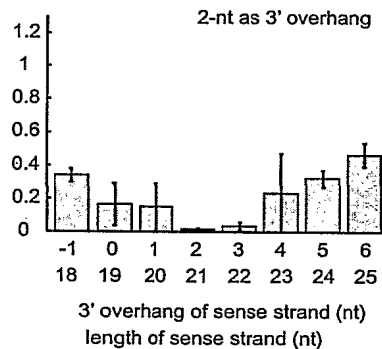
D

```
5' CGUACGCGGAAUACUUCG
   AGUGCAUGCGCCUUAUGAAGC 5'

5' CGUACGCGGAAUACUUCGA
   AGUGCAUGCGCCUUAUGAAGC 5'

5' CGUACGCGGAAUACUUCGAA
   AGUGCAUGCGCCUUAUGAAGC 5'

5' CGUACGCGGAAUACUUCGAAA
   AGUGCAUGCGCCUUAUGAAGC 5'

5' CGUACGCGGAAUACUUCGAAAU
   AGUGCAUGCGCCUUAUGAAGC 5'

5' CGUACGCGGAAUACUUCGAAAUG
   AGUGCAUGCGCCUUAUGAAGC 5'

5' CGUACGCGGAAUACUUCGAAAUGU
   AGUGCAUGCGCCUUAUGAAGC 5'

5' CGUACGCGGAAUACUUCGAAAUGUC
   AGUGCAUGCGCCUUAUGAAGC 5'
```
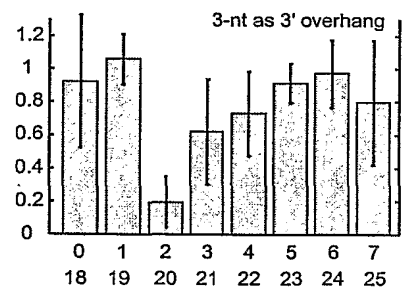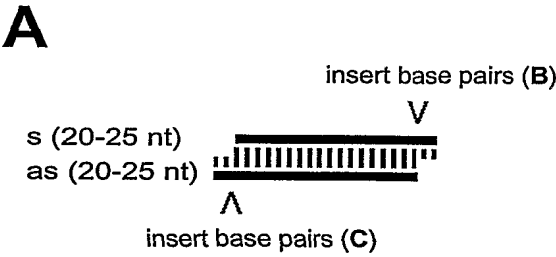
1-nt as 3' overhang

norm. Pp-luc/Rr-luc

-2 -1   0   1   2   3   4   5
18 19 20 21 22 23 24 25

2-nt as 3' overhang

-1   0   1   2   3   4   5   6
18 19 20 21 22 23 24 25

3' overhang of sense strand (nt)
length of sense strand (nt)

3-nt as 3' overhang

0   1   2   3   4   5   6   7
18 19 20 21 22 23 24 25

Fig. 13

A

s (20-25 nt)
as (20-25 nt)

insert base pairs (B)

insert base pairs (C)

B

5' CGUACGCGGAAUACUUCGAA
GUGCAUGCGCCUUAUGAAGC 5'

5' CGUACGCGGAAUACUUCGAAA
GUGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGAAAU
GUGCAUGCGCCUUAUGAAGCUU 5'

5' CGUACGCGGAAUACUUCGAAAUG
GUGCAUGCGCCUUAUGAAGCUUU 5'

5' CGUACGCGGAAUACUUCGAAAUGU
GUGCAUGCGCCUUAUGAAGCUUUA 5'

5' CGUACGCGGAAUACUUCGAAAUGUC
GUGCAUGCGCCUUAUGAAGCUUUAC 5'
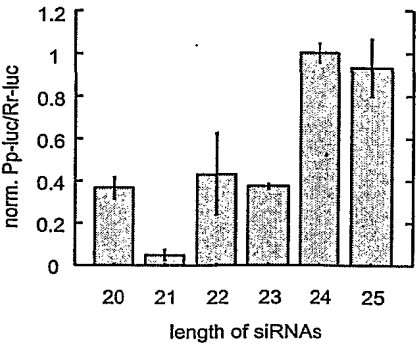
C

5' GUACGCGGAAUACUUCGAAA
UGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGAAA
GUGCAUGCGCCUUAUGAAGCU 5'

5' ACGUACGCGGAAUACUUCGAAA
AGUGCAUGCGCCUUAUGAAGCU 5'

5' CACGUACGCGGAAUACUUCGAAA
UAGUGCAUGCGCCUUAUGAAGCU 5'
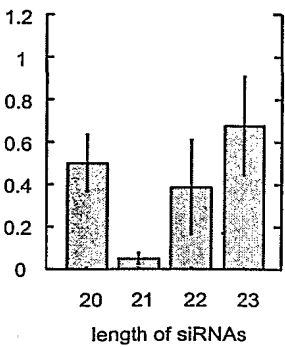
Fig. 14

s   5′ CGUACGCGGAAUACUUCGAAA
as   GUGCAUGCGCCUUAUGAAGCU 5′
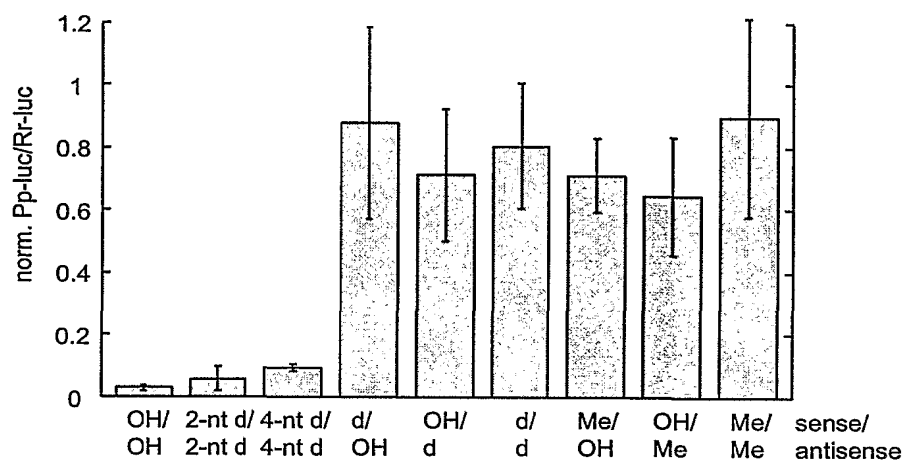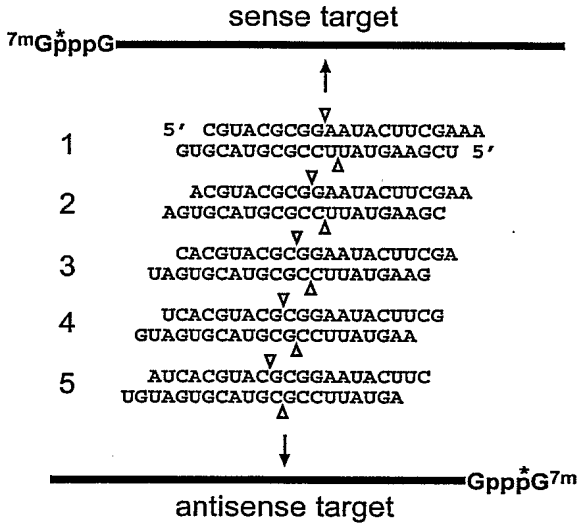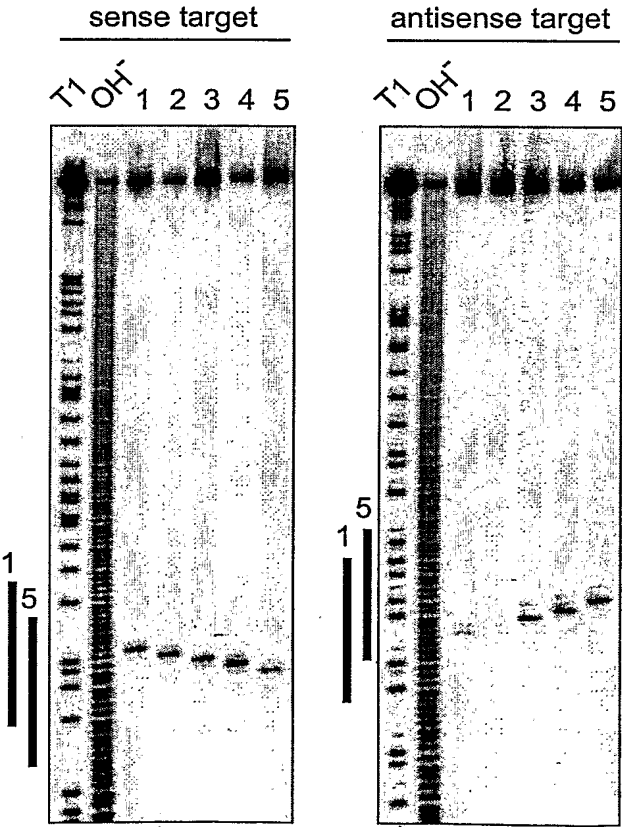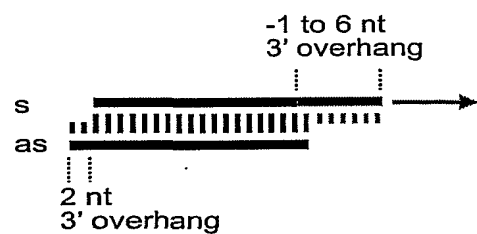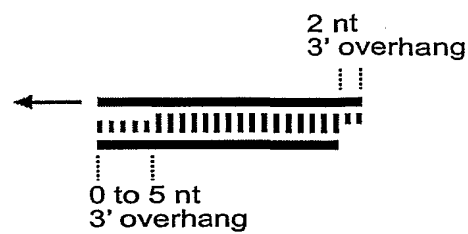
A

Fig. 15

```
                          sense target
7mGpppG ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━
                                    ↑
                                    ▽
        1    5' CGUACGCGGAAUACUUCGAAA
                GUGCAUGCGCCUUAUGAAGCU 5'
                                  △
                             ▽
        2       ACGUACGCGGAAUACUUCGAA
                AGUGCAUGCGCCUUAUGAAGC
                                 △
                           ▽
        3       CACGUACGCGGAAUACUUCGA
                UAGUGCAUGCGCCUUAUGAAG
                                △
                          ▽
        4       UCACGUACGCGGAAUACUUCG
                GUAGUGCAUGCGCCUUAUGAA
                               △
                         ▽
        5       AUCACGUACGCGGAAUACUUC
                UGUAGUGCAUGCGCCUUAUGA
                              △
                                    ↓
        ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ GpppG7m
                          antisense target
```

B

Fig.16

A

-1 to 6 nt
3' overhang

s

as

2 nt
3' overhang

B

2 nt
3' overhang

0 to 5 nt
3' overhang

C

5' CGUACGCGGAAUACUUCG
GUGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGA
GUGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGAA
GUGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGAAA
GUGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGAAAU
GUGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGAAAUG
GUGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGAAAUGU
GUGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGAAAUGUC
GUGCAUGCGCCUUAUGAAGCU 5'
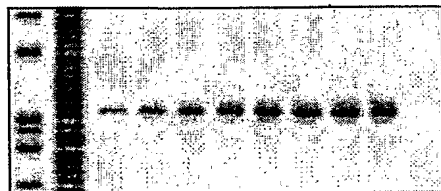
D

5' ACGCGGAAUACUUCGAAA
GUGCAUGCGCCUUAUGAAGCU 5'

5' UACGCGGAAUACUUCGAAA
GUGCAUGCGCCUUAUGAAGCU 5'

5' GUACGCGGAAUACUUCGAAA
GUGCAUGCGCCUUAUGAAGCU 5'

5' CGUACGCGGAAUACUUCGAAA
GUGCAUGCGCCUUAUGAAGCU 5'

5' ACGUACGCGGAAUACUUCGAAA
GUGCAUGCGCCUUAUGAAGCU 5'

5' CACGUACGCGGAAUACUUCGAAA
GUGCAUGCGCCUUAUGAAGCU 5'

T1 OH 18 19 20 21 22 23 24 25 -

sense
target

T1 OH 18 19 20 21 22 23

antisense
target

Fig. 17

s/as overhanging nucleotides
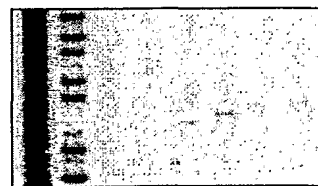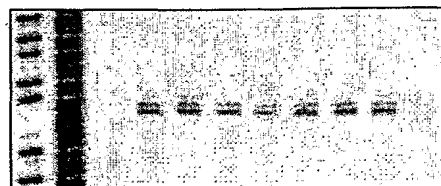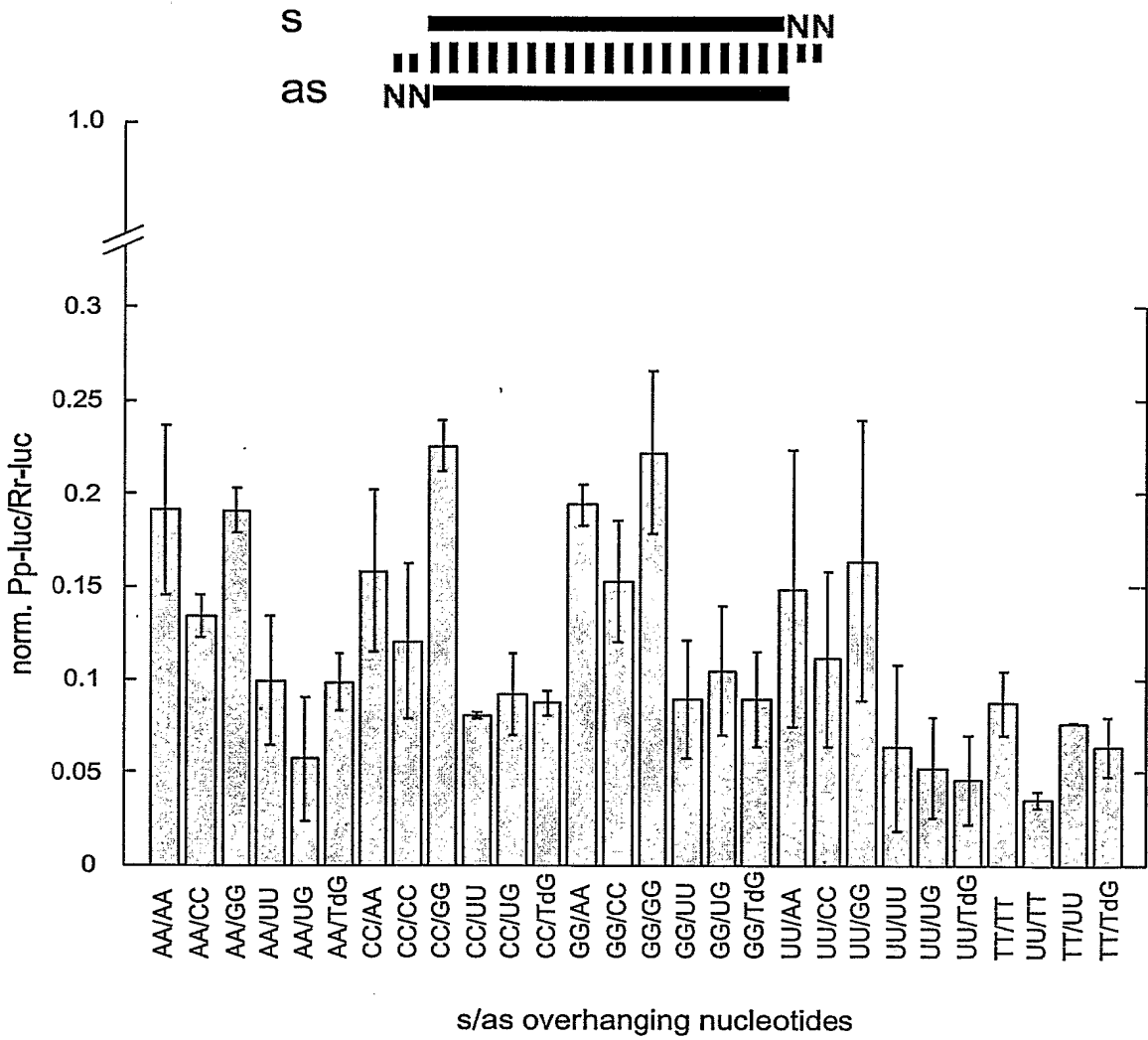
Fig. 18

```
ref   5'  CGUACGCGGAAUACUUCGATT
          TTGCAUGCGCCUUAUGAAGCU  5'

 1    5'  AUGCCGCGGAAUACUUCGATT
          TTUACGGCGCCUUAUGAAGCU  5'

 2    5'  CGUAGCGCGAAUACUUCGATT
          TTGCAUCGCGCUUAUGAAGCU  5'

 3    5'  CGUACGCGAGUAACUUCGATT
          TTGCAUGCGCUCAUUGAAGCU  5'

 4    5'  CGUACGCGGAAUUUCACGATT
          TTGCAUGCGCCUUAAAGUGCU  5'

 5    5'  CGUACGCGGAAUACUUAGGTT
          TTGCAUGCGCCUUAUGAAUCG  5'

 6    5'  CGUACGCGGUAUACUUCGATT
          TTGCAUGCGCCAUAUGAAGCU  5'

 7    5'  CGUACGCGGAUUACUUCGATT
          TTGCAUGCGCCUAAUGAAGCU  5'
```

Figure 19

A

```
5'  CGUACGCGGAAUACUUCGAA
     GUGCAUGCGCCUUAUGAAGC  5'

5'  CGUACGCGGAAUACUUCGAAA
     GUGCAUGCGCCUUAUGAAGCU  5'

5'  CGUACGCGGAAUACUUCGATT
     TTGCAUGCGCCUUAUGAAGCU  5'

5'  CGUACGCGGAAUACUUCGAAAU
     GUGCAUGCGCCUUAUGAAGCUU  5'

5'  CGUACGCGGAAUACUUCGAAAUG
     GUGCAUGCGCCUUAUGAAGCUUU  5'

5'  CGUACGCGGAAUACUUCGAAAUGU
     GUGCAUGCGCCUUAUGAAGCUUUA  5'

5'  CGUACGCGGAAUACUUCGAAAUGUC
     GUGCAUGCGCCUUAUGAAGCUUUAC  5'
```

B

```
5'  GUACGCGGAAUACUUCGAAA
     UGCAUGCGCCUUAUGAAGCU  5'

5'  CGUACGCGGAAUACUUCGAAA
     GUGCAUGCGCCUUAUGAAGCU  5'

5'  ACGUACGCGGAAUACUUCGAAA
     AGUGCAUGCGCCUUAUGAAGCU  5'

5'  CACGUACGCGGAAUACUUCGAAA
     UAGUGCAUGCGCCUUAUGAAGCU  5'
```

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(71) Applicant (for all designated States except US): RI-BOZYME PHARMACEUTICALS, INC [US/US]; 2950 Wilderness Place, Boulder, CO 803013 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): BEIGELMAN, Leonid [US/US]; 5530 Colt Drive, Longmont, CO 80503 (US). MATULIC-ADAMIC, Jasenka [US/US]; 760 S. 42nd Street, Boulder, CO 80303 (US). VARGEESE, Chandra [US/US]; 3295 E. 117th Avenue, Thornton, CO 80233 (US). KARPEISKY, Alexander [US/US]; 420 Vernier Avenue, Lafayette, CO 80026 (US). BLATT, Lawrence [US/US]; 2176 Riverside Lane, Boulder, CO

80304 (US). SHAFFER, Christopher [US/US]; 9141 Ute Highway, Longmont, CO 80503 (US).

(74) Agent: DOCTER, Stephen, H.; McDonnell Boehnen Hulbert & Berghoff, 300 s. Wacker Drive, Suite 3200, Chicago, IL60606 (US).

(54) Title: CONJUGATES AND COMPOSITIONS FOR CELLULAR DELIVERY

(57) Abstract: This invention features conjugates, degradable linkers, compositions, methods of synthesis, and applications thereof, including folate, galactose, galactosamine, N-acetyl galactosamine, PEG, phospholipid, peptide and human serum albumin (HAS) derived conjugates of biologically active compounds, including antibodies, antivirals, chemotherapeutics, peptides, proteins, hormones nucleosides, nucleotides, non-nucleosides, and nucleic acids including enzymatic nucleic acids, DNAzymes, allozymes, antisense, dsRNA, siRNA, triplex oligonucleotides, 2,5-A chimeras, decoys and aptamers.

CONJUGATES AND COMPOSITIONS FOR CELLULAR DELIVERY

Background of the Invention

This patent application claims priority from Adamic *et al.*, USSN (60/292,217), filed May 18, 2001, and Adamic *et al.*, USSN (60/362,016) filed March 6, 2002 both entitled 'CONJUGATES AND COMPOSITIONS FOR CELLULAR DELIVERY'. This patent application also claims priority from Vargeese *et al.*, USSN (60/306,883), filed July 20, 2001 entitled "CONJUGATES AND COMPOSITIONS FOR TRANSPORT ACROSS CELLULAR MEMBRANES" and Vargeese *et al.*, USSN (60/311,865), filed August 13, 2001, entitled "CONJUGATES AND COMPOSITIONS FOR CELLULAR DELIVERY". These applications are hereby incorporated by reference herein in their entirety including the drawings.

The present invention relates to conjugates, compositions, methods of synthesis, and applications thereof. The discussion is provided only for understanding of the invention that follows. This summary is not an admission that any of the work described below is prior art to the claimed invention.

The cellular delivery of various therapeutic compounds, such as antiviral and chemotherapeutic agents, is usually compromised by two limitations. First the selectivity of chemotherapeutic agents is often low, resulting in high toxicity to normal tissues. Secondly, the trafficking of many compounds into living cells is highly restricted by the complex membrane systems of the cell. Specific transporters allow the selective entry of nutrients or regulatory molecules, while excluding most exogenous molecules such as nucleic acids and proteins. Various strategies can be used to improve transport of compounds into cells, including the use of lipid carriers and various conjugate systems. Conjugates are often selected based on the ability of certain molecules to be selectively transported into specific cells, for example via receptor mediated endocytosis. By attaching a compound of interest to molecules that are actively transported across the cellular membranes, the effective transfer of that compound into cells or specific cellular organelles can be realized. Alternately, molecules that are able to penetrate cellular membranes without active transport mechanisms, for example, various lipophilic molecules, can be used to deliver compounds of interest. Examples of molecules that can

be utilized as conjugates include but are not limited to peptides, hormones, fatty acids, vitamins, flavonoids, sugars, reporter molecules, reporter enzymes, chelators, porphyrins, intercalcators, and other molecules that are capable of penetrating cellular membranes, either by active transport or passive transport.

5   The delivery of compounds to specific cell types, for example, cancer cells, can be accomplished by utilizing receptors associated with specific cell types. Particular receptors are overexpressed in certain cancerous cells, including the high affinity folic acid receptor. For example, the high affinity folate receptor is a tumor marker that is overexpressed in a variety of neoplastic tissues, including breast, ovarian, cervical,

10   colorectal, renal, and nasoparyngeal tumors, but is expressed to a very limited extent in normal tissues. The use of folic acid based conjugates to transport exogenous compounds across cell membranes can provide a targeted delivery approach to the treatment and diagnosis of disease and can provide a reduction in the required dose of therapeutic compounds. Furthermore, therapeutic bioavialability, pharmacodynamics, and

15   pharmacokinetic parameters can be modulated through the use of bioconjugates, including folate bioconjugates. Godwin *et al.*, 1972, *J. Biol. Chem.*, 247, 2266-2271, report the synthesis of biologically active pteroyloligo-L-glutamates. Habus *et al.*, 1998, *Bioconjugate Chem.*, 9, 283-291, describe a method for the solid phase synthesis of certain oligonucleotide-folate conjugates. Cook, US Patent No. 6,721,208, describes

20   certain oligonucleotides modified with specific conjugate groups. The use of biotin and folate conjugates to enhance transmembrane transport of exogenous molecules, including specific oligonucleotides has been reported by Low *et al.*, US Patent Nos. 5,416,016, 5,108,921, and International PCT publication No. WO 90/12096. Manoharan *et al.*, International PCT publication No. WO 99/66063 describe certain folate conjugates,

25   including specific nucleic acid folate conjugates with a phosphoramidite moiety attached to the nucleic acid component of the conjugate, and methods for the synthesis of these folate conjugates. Nomura *et al.*, 2000, *J. Org. Chem.*, 65, 5016-5021, describe the synthesis of an intermediate, alpha-[2-(trimethylsilyl)ethoxycarbonl]folic acid, useful in the synthesis of ceratin types of folate-nucleoside conjugates. Guzaev *et al.*, US

30   6,335,434, describes the synthesis of certain folate oligonucleotide conjugates.

The delivery of compounds to other cell types can be accomplished by utilizing receptors associated with a certain type of cell, such as hepatocytes. For example, drug

delivery systems utilizing receptor-mediated endocytosis have been employed to achieve drug targeting as well as drug-uptake enhancement. The asialoglycoprotein receptor (ASGPr) (see for example Wu and Wu, 1987, *J. Biol. Chem.* 262, 4429-4432) is unique to hepatocytes and binds branched galactose-terminal glycoproteins, such as asialoorosomucoid (ASOR). Binding of such glycoproteins or synthetic glycoconjugates to the receptor takes place with an affinity that strongly depends on the degree of branching of the oligosaccharide chain, for example, triatennary structures are bound with greater affinity than biatenarry or monoatennary chains (Baenziger and Fiete, 1980, *Cell,* 22, 611-620; Connolly *et al.,* 1982, *J. Biol. Chem.,* 257, 939-945). Lee and Lee, 1987, *Glycoconjugate J.,* 4, 317-328, obtained this high specificity through the use of N-acetyl-D-galactosamine as the carbohydrate moiety, which has higher affinity for the receptor, compared to galactose. This "clustering effect" has also been described for the binding and uptake of mannosyl-terminating glycoproteins or glycoconjugates (Ponpipom *et al.,* 1981, *J. Med. Chem.,* 24, 1388-1395). The use of galactose and galactosamine based conjugates to transport exogenous compounds across cell membranes can provide a targeted delivery approach to the treatment of liver disease such as HBV and HCV infection or hepatocellular carcinoma. The use of bioconjugates can also provide a reduction in the required dose of therapeutic compounds required for treatment. Furthermore, therapeutic bioavialability, pharmacodynamics, and pharmacokinetic parameters can be modulated through the use of bioconjugates.

A number of peptide based cellular transporters have been developed by several research groups. These peptides are capable of crossing cellular membranes *in vitro* and *in vivo* with high efficiency. Examples of such fusogenic peptides include a 16-amino acid fragment of the homeodomain of ANTENNAPEDIA, a Drosophila transcription factor (Wang *et al.,* 1995, *PNAS USA.,* 92, 3318-3322); a 17-mer fragment representing the hydrophobic region of the signal sequence of Kaposi fibroblast growth factor with or without NLS domain (Antopolsky *et al.,* 1999, *Bioconj. Chem.,* 10, 598-606); a 17-mer signal peptide sequence of caiman crocodylus Ig(5) light chain (Chaloin *et al.,* 1997, *Biochem. Biophys. Res. Comm.,* 243, 601-608); a 17-amino acid fusion sequence of HIV envelope glycoprotein gp4114, (Morris *et al.,*1997, *Nucleic Acids Res.,* 25, 2730-2736); the HIV-1 Tat49-57 fragment (Schwarze *et al.,* 1999, *Science,* 285, 1569-1572); a transportan A - achimeric 27-mer consisting of N-terminal fragment of neuropeptide

4

galanine and membrane interacting wasp venom peptide mastoporan (Lindgren *et al.*, 2000, *Bioconjugate Chem.*, 11, 619-626); and a 24-mer derived from influenza virus hemagglutinin envelop glycoprotein (Bongartz *et al.*, 1994, *Nucleic Acids Res.*, 22, 4681-4688).

5   These peptides were successfully used as part of an antisense oligonucleotide-peptide conjugate for cell culture transfection without lipids. In a number of cases, such conjugates demonstrated better cell culture efficacy then parent oligonucleotides transfected using lipid delivery. In addition, use of phage display techniques has identified several organ targeting and tumor targeting peptides *in vivo* (Ruoslahti, 1996, *Ann. Rev.*

10   *Cell Dev. Biol.*, 12, 697-715). Conjugation of tumor targeting peptides to doxorubicin has been shown to significantly improve the toxicity profile and has demonstrated enhanced efficacy of doxorubicin in the *in vivo* murine cancer model MDA-MB-435 breast carcinoma (Arap *et al.*, 1998, *Science*, 279, 377-380).

Hudson *et al.*, 1999, *Int. J. Pharm.*, 182, 49-58, describes the cellular delivery of

15   specific hammerhead ribozymes conjugated to a transferrin receptor antibody. Janjic *et al.*, US 6,168,778, describes specific VEGF nucleic acid ligand complexes for targeted drug delivery. Bonora *et al.*, 1999, *Nucleosides Nucleotides*, 18, 1723-1725, describes the biological properties of specific antisense oligonucleotides conjugated to certain polyethylene glycols. Davis and Bishop, International PCT publication No. WO

20   99/17120 and Jaeschke *et al.*, 1993, *Tetrahedron Lett.*, 34, 301-4 describe specific methods of preparing polyethylene glycol conjugates. Tullis, International PCT Publication No. WO 88/09810; Jaschke, 1997, *ACS Sympl Ser.*, 680, 265-283; Jaschke *et al.*, 1994, *Nucleic Acids Res.*, 22, 4810-17; Efimov *et al.*, 1993, *Bioorg. Khim.*, 19, 800-4; and Bonora *et al.*, 1997, *Bioconjugate Chem.*, 8, 793-797, describe specific

25   oligonucleotide polyethylene glycol conjugates. Manoharan, International PCT Publication No. WO 00/76554, describes the preparation of specific ligand-conjugated oligodeoxyribonucleotides with certain cellular, serum, or vascular proteins. Defrancq and Lhomme, 2001, *Bioorg Med Chem Lett.*, 11, 931-933; Cebon *et al.*, 2000, Aust. J. Chem., 53, 333-339; and Salo *et al.*, 1999, *Bioconjugate Chem.*, 10, 815-823 describe

30   specific aminooxy peptide oligonucleotide conjugates.

Summary of the Invention

The present invention features compositions and conjugates to facilitate delivery of molecules into a biological system, such as cells. The conjugates provided by the instant
5   invention can impart therapeutic activity by transferring therapeutic compounds across cellular membranes. The present invention encompasses the design and synthesis of novel agents for the delivery of molecules, including but not limited to small molecules, lipids, nucleosides, nucleotides, nucleic acids, antibodies, toxins, negatively charged polymers and other polymers, for example proteins, peptides, hormones, carbohydrates, or
10   polyamines, across cellular membranes. In general, the transporters described are designed to be used either individually or as part of a multi-component system, with or without degradable linkers. The compounds of the invention generally shown in the Formulae below are expected to improve delivery of molecules into a number of cell types originating from different tissues, in the presence or absence of serum.

15   The present invention features a compound having the Formula 1:

1

wherein each R_1, R_3, R_4 ,R_5, R_6, R_7 and R_8 is independently hydrogen, alkyl , substituted alkyl, aryl, substituted aryl, or a protecting group, each "n" is independently an
20   integer from 0 to about 200, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and R_2 is a phosphorus containing group, nucleoside, nucleotide, small molecule, nucleic acid, or a solid support comprising a linker.

The present invention features a compound having the Formula 2:

25   2

SUBSTITUTE SHEET (RULE 26)

wherein each R_3, R_4 ,R_5, R_6 and R_7 is independently hydrogen, alkyl , substituted alkyl, aryl, substituted aryl, or a protecting group, each "n" is independently an integer from 0 to about 200, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and R_2 is a phosphorus containing group, nucleoside, nucleotide, small molecule, nucleic acid, or a solid support comprising a linker.

The present invention features a compound having the Formula 3:

3

wherein each R_1, R_3, R_4 ,R_5 R_6 and R_7 is independently hydrogen, alkyl , substituted alkyl, aryl, substituted aryl, or a protecting group, each "n" is independently an integer from 0 to about 200, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and R_2 is a phosphorus containing group, nucleoside, nucleotide, small molecule, or nucleic acid.

The present invention features a compound having the Formula 4:
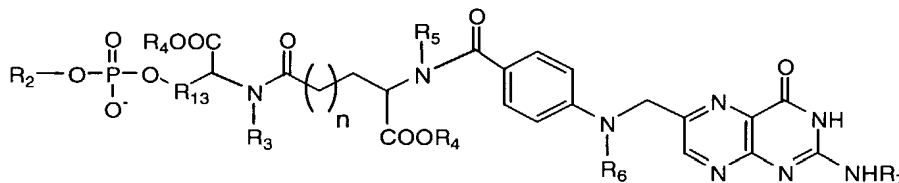
4

wherein each R_3, R_4 ,R_5, R_6 and R_7 is independently hydrogen, alkyl , substituted alkyl, aryl, substituted aryl, or a protecting group, each "n" is independently an integer from 0 to about 200, R_2 is a phosphorus containing group, nucleoside, nucleotide, small molecule, nucleic acid, or a solid support comprising a linker, and R_{13} is an amino acid side chain.

The present invention features a compound having the Formula 5:
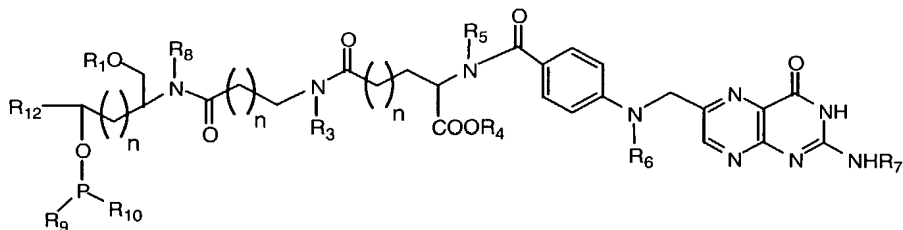
5

wherein each R_1 and R_4 is independently a protecting group or hydrogen, each R_3, R_5, R_6, R_7 and R_8 is independently hydrogen, alkyl or nitrogen protecting group, each "n" is independently an integer from 0 to about 200, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and each R_9 and R_{10} is independently a nitrogen containing group, cyanoalkoxy, alkoxy, aryloxy, or alkyl group.

The present invention features a compound having the Formula 6:

6

wherein each R_4, R_5, R_6 and R_7 is independently hydrogen, alkyl, substituted alkyl, aryl, substituted aryl, or a protecting group, R_2 is a phosphorus containing group, nucleoside, nucleotide, small molecule, nucleic acid, or a solid support comprising a linker, each "n" is independently an integer from 0 to about 200, and L is a degradable linker.

The present invention features a compound having the Formula 7:
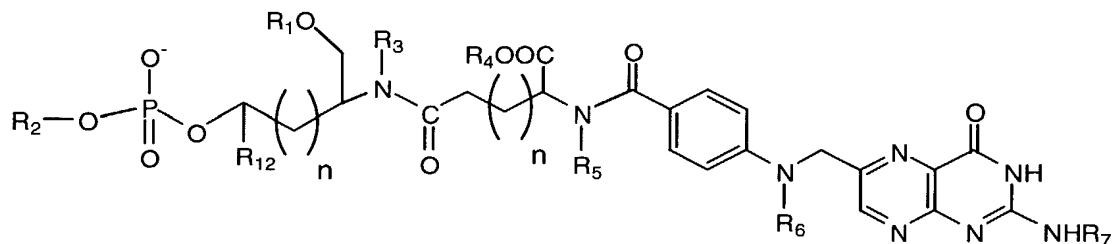
7

wherein each R_1, R_3, R_4, R_5, R_6 and R_7 is independently hydrogen, alkyl, substituted alkyl, aryl, substituted aryl, or a protecting group, each "n" is independently an integer from 0 to about 200, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and R_2 is a phosphorus containing group, nucleoside, nucleotide, small molecule, nucleic acid, or a solid support comprising a linker.

The present invention features a compound having the Formula 8:

8

wherein each R_1 and R_4 is independently a protecting group or hydrogen, each R_3, R_5, R_6 and R_7 is independently hydrogen, alkyl or nitrogen protecting group, each "n" is independently an integer from 0 to about 200, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and each R_9 and R_{10} is independently a nitrogen containing group, cyanoalkoxy, alkoxy, aryloxy, or alkyl group.

The present invention features a method for synthesizing a compound having Formula 5:

5

wherein each R_1 and R_4 is independently a protecting group or hydrogen, each R_3, R_5, R_6 and R_7 is independently hydrogen, alkyl or nitrogen protecting group, each "n" is independently an integer from 0 to about 200, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and each R_9 and R_{10} is independently a nitrogen containing group, cyanoalkoxy, alkoxy, aryloxy, or alkyl group, comprising: coupling a bis-hydroxy aminoalkyl derivative, for example D-threoninol, with a N-protected aminoalkanoic acid to yield a compound of Formula 9;

9
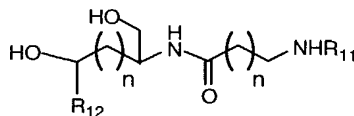
9

wherein R_{11} is an amino protecting group, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and each "n" is independently an integer from 0 to about 200; introducing primary hydroxy protection R_1 followed by amino deprotection of R_{11} to yield a compound of Formula 10;
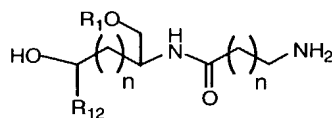
10

wherein R_1 is a protecting group, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and each "n" is independently an integer from 0 to about 200; coupling the deprotected amine of Formula 10 with a protected amino acid, for example glutamic acid, to yield a compound of Formula 11;
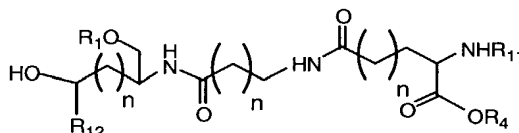
11

wherein each R_1 and R_4 is independently a protecting group or hydrogen, each "n" is independently an integer from 0 to about 200, R_{11} is an amino protecting group, and R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl; deprotecting the amine R_{11} of the conjugated glutamic acid of Formula XI to yield a compound of Formula 12;
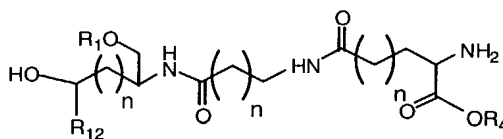
12

wherein each R_1 and R_4 is independently a protecting group or hydrogen, each "n" is independently an integer from 0 to about 200, R_{11} is an amino protecting group, and

R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl; coupling the deprotected amine of Formula 12 with an amino protected pteroic acid to yield a compound of Formula 13;
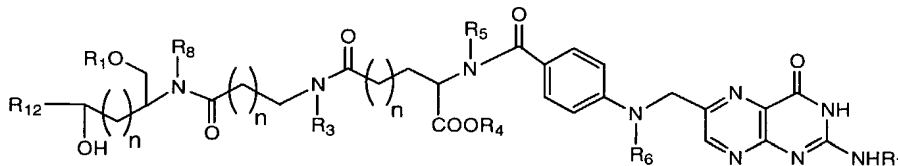
13

5

wherein each R_1 and R_4 is independently a protecting group or hydrogen, each R_3, R_5, R_6 and R_7 is independently hydrogen, alkyl or nitrogen protecting group, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and each "n" is independently an integer from 0 to about 200; and introducing a phosphorus containing group at the secondary hydroxyl of Formula 13 to yield a compound of Formula 5.

10

The present invention features a method for synthesizing a compound having Formula 8:
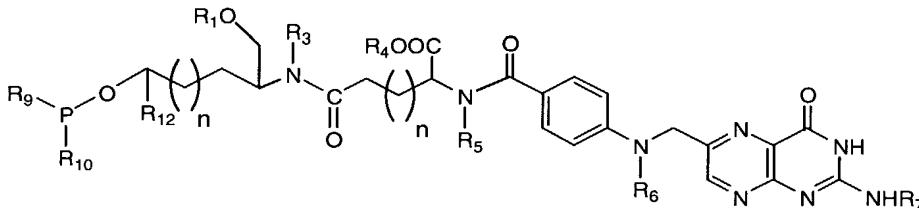
8

wherein each R_1 and R_4 is independently a protecting group or hydrogen, each R_3, R_5, R_6 and R_7 is independently hydrogen, alkyl or nitrogen protecting group, each "n" is independently an integer from 0 to about 200, each R_9 and R_{10} is independently a nitrogen containing group, cyanoalkoxy, alkoxy, aryloxy, or alkyl group, and R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, comprising; coupling a bis-hydroxy aminoalkyl derivative, for example D-threoninol, with a protected amino acid, for example glutamic acid, to yield a compound of Formula 14;

15

20

14
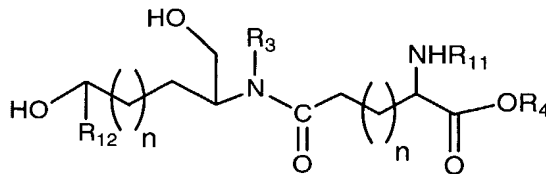
wherein R_{11} is an amino protecting group, each "n" is independently an integer from 0 to about 200, R_4 is independently a protecting group, and R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl; introducing primary hydroxy protection R_1 followed by amino deprotection of R_{11} of Formula 14 to yield a compound of Formula 15;

15
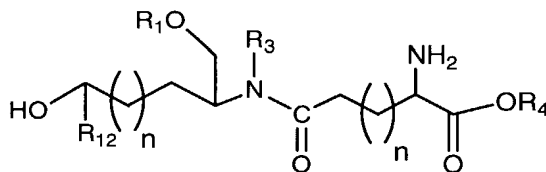
wherein each R_1 and R_4 is independently a protecting group or hydrogen, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and each "n" is independently an integer from 0 to about 200; coupling the deprotected amine of Formula 15 with an amino protected pteroic acid to yield a compound of Formula 16;
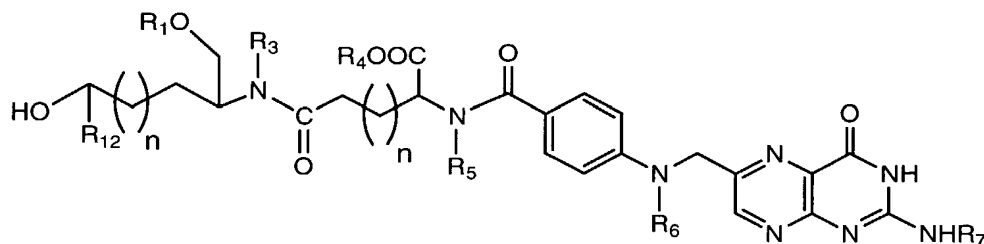
16

wherein each R_1 and R_4 is independently a protecting group or hydrogen, each R_3, R_5, R_6 and R_7 is independently hydrogen, alkyl or nitrogen protecting group, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and each "n" is independently an integer from 0 to about 200; and introducing a phosphorus containing group at the secondary hydroxyl of Formula 16 to yield a compound of Formula 8.

In one embodiment, R_2 of a compound of the invention comprises a phosphorus containing group.

In another embodiment, R_2 of a compound of the invention comprises a nucleoside, for example, a nucleoside with beneficial activity such as anticancer or antiviral activity.

In yet another embodiment, R_2 of a compound of the invention comprises a nucleotide, for example, a nucleotide with beneficial activity such as anticancer or antiviral activity.

In a further embodiment, R_2 of a compound of the invention comprises a small molecule, for example, a small molecule with beneficial activity such as anticancer or antiviral activity.

In another embodiment, R_2 of a compound of the invention comprises a nucleic acid, for example, a nucleic acid with beneficial activity such as anticancer or antiviral activity..

In one embodiment, R_2 of a compound of the invention comprises a solid support comprising a linker.

In another embodiment, a nucleoside (R_2) of the invention comprises a nucleoside with anticancer activity.

In another embodiment, a nucleoside (R_2) of the invention comprises a nucleoside with antiviral activity.

In another embodiment, the nucleoside (R_2) of the invention comprises fludarabine, lamivudine (3TC), 5-fluro uridine, AZT, ara-adenosine, ara-adenosine monophosphate, a dideoxy nucleoside analog, carbodeoxyguanosine, ribavirin, fialuridine, lobucavir, a pyrophosphate nucleoside analog, an acyclic nucleoside analog, acyclovir, gangciclovir, penciclovir, famciclovir, an L-nucleoside analog, FTC, L-FMAU, L-ddC, L-FddC, L-d4C, L-Fd4C, an L-dideoxypurine nucleoside analog, cytallene, bis-POM PMEA (GS-840), BMS-200,475, carbovir or abacavir.

In one embodiment, R_{13} of a compound of the invention comprises an alkylamino or an alkoxy group, for example, $-CH_2O-$ or $-CH(CH_2)CH_2O-$.

In another embodiment, R_{12} of a compound of the invention is an alkylhyrdroxyl, for example, $-(CH_2)_nOH$, where n comprises an integer from about 1 to about 10.

In another embodiment, L of Formula 6 of the invention comprises serine, threonine, or a photolabile linkage.

In one embodiment, R_9 of a compound of the invention comprises a phosphorus protecting group, for example -OCH_2CH_2CN (oxyethylcyano).

In one embodiment, R_{10} of a compound of the invention comprises a nitrogen containing group, for example, –$N(R_{14})$ wherein R_{14} is a straight or branched chain alkyl having from about 1 to about 10 carbons.

In another embodiment, R_{10} of a compound of the invention comprises a heterocycloalkyl or heterocycloalkenyl ring containing from about 4 to about 7 atoms, and having from about 1 to about 3 heteroatoms comprising oxygen, nitrogen, or sulfur.

In another embodiment, R_1 of a compound of the invention comprises an acid labile protecting group, such as a trityl or substituted trityl group, for example, a dimethoxytrityl or mono-methoxytrityl group.

In another embodiment, R_4 of a compound of the invention comprises a *tert*-butyl, Fm (fluorenyl-methoxy), or allyl group.

In one embodiment, R_6 of a compound of the invention comprises a TFA (trifluoracetyl) group.

In another embodiment, R_3, R_5 R_7 and R_8 of a compound of the invention are independently hydrogen.

In one embodiment, R_7 of a compound of the invention is independently isobutyryl, dimethylformamide, or hydrogen.

In another embodiment, R_{12} of a compound of the invention comprises a methyl group or ethyl group.

In one embodiment, a nucleic acid of the invention comprises an enzymatic nucleic acid, for example a hammerhead, Inozyme, DNAzyme, G-cleaver, Zinzyme, Amberzyme, or allozyme.

In another embodiment, a nucleic acid of the invention comprises an antisense nucleic acid, 2-5A nucleic acid chimera, or decoy nucleic acid.

In another embodiment, the solid support having a linker of the invention comprises a structure of Formula 17:
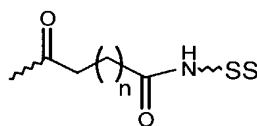
17

wherein SS is a solid support, and each "n" is independently an integer from about 1 to about 200.

In another embodiment, the solid support of the instant invention is controlled pore glass (CPG) or polystyrene, and can be used in the synthesis of a nucleic acid.

In one embodiment, the invention features a pharmaceutical composition comprising a compound of the invention and a pharmaceutically acceptable carrier.

In another embodiment, the invention features a method of treating a cancer patient, comprising contacting cells of the patient with a pharmaceutical composition of the invention under conditions suitable for the treatment. This treatment can comprise the use of one or more other drug therapies under conditions suitable for the treatment. The cancers contemplated by the instant invention include but are not limited to breast cancer, lung cancer, colorectal cancer, brain cancer, esophageal cancer, stomach cancer, bladder cancer, pancreatic cancer, cervical cancer, head and neck cancer, ovarian cancer, melanoma, lymphoma, glioma, or multidrug resistant cancers.

In one embodiment, the invention features a method of treating a patient infected with a virus, comprising contacting cells of the patient with a pharmaceutical composition of the invention, under conditions suitable for the treatment. This treatment can comprise the use of one or more other drug therapies under conditions suitable for the treatment. The viruses contemplated by the instant invention include but are not limited to HIV, HBV, HCV, CMV, RSV, HSV, poliovirus, influenza, rhinovirus, west nile virus, Ebola virus, foot and mouth virus, and papilloma virus.

In one embodiment, the invention features a kit for detecting the presence of a nucleic acid molecule or other target molecule in a sample, for example, a gene in a cancer cell, comprising a compound of the instant invention.

In one embodiment, the invention features a kit for detecting the presence of a nucleic acid molecule, or other target molecule in a sample, for example, a gene in a virus-infected cell, comprising a compound of the instant invention.

In another embodiment, the invention features a compound of the instant invention comprising a modified phosphate group, for example, a phosphoramidite, phosphodiester, phosphoramidate, phosphorothioate, phosphorodithioate, alkylphosphonate, arylphosphonate, monophosphate, diphosphate, triphosphate, or pyrophosphate.

In one embodiment, the invention features a method for synthesizing a compound having Formula 18:

18
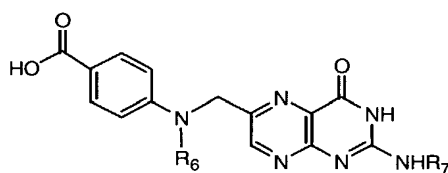
wherein each R_6 and R_7 is independently hydrogen, alkyl or nitrogen protecting group, comprising: reacting folic acid with a carboxypeptidase to yield a compound of Formula 19;
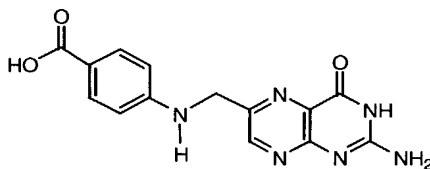
19

introducing a protecting group R_6 on the secondary amine of Formula 19 to yield a compound of Formula 20;

20
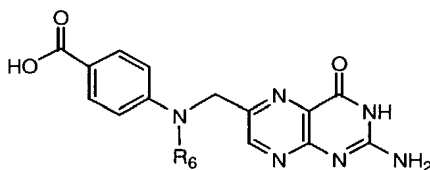
wherein R_6 is a nitrogen protecting group; and introducing a protecting group R_7 on the primary amine of Formula 20 to yield a compound of Formula 18.

In another embodiment, the amino protected pteroic acid of the invention is a compound of Formula 18.

16

In one embodiment, the invention encompasses a compound of Formula 1 having Formula 21:
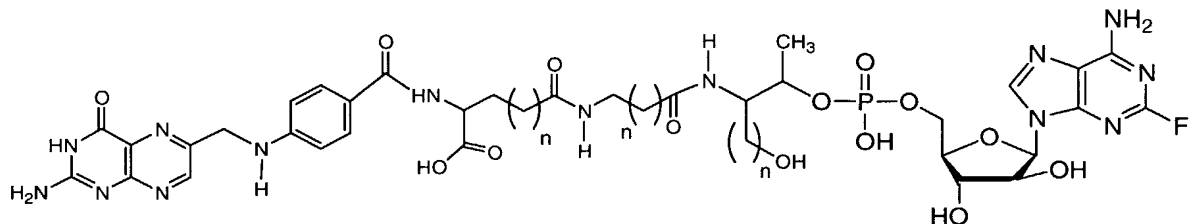
21

5   wherein each "n" is independently an integer from 0 to about 200.

In another embodiment, the invention encompasses a compound of Formula 7 having Formula 22:
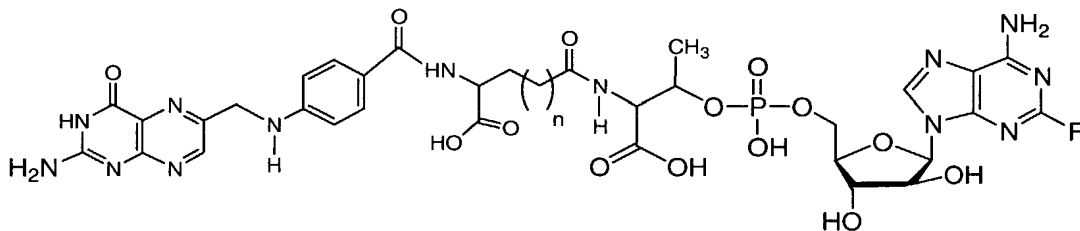
22

10   wherein each "n" is independently an integer from 0 to about 200.

In another embodiment, the invention encompasses a compound of Formula 4 having Formula 23:
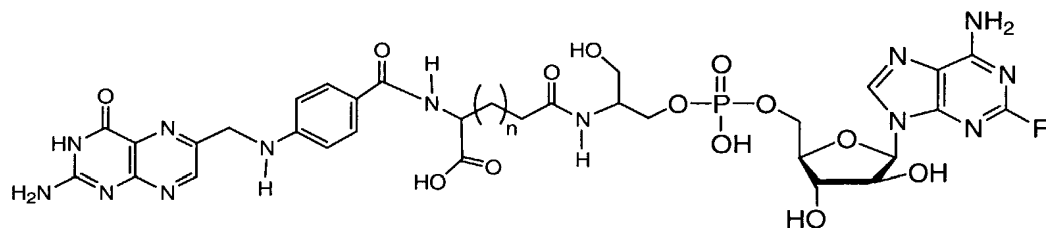
23

15   wherein "n" is an integer from 0 to about 200.

In another embodiment, the invention encompasses a compound of Formula 4 having Formula 24:
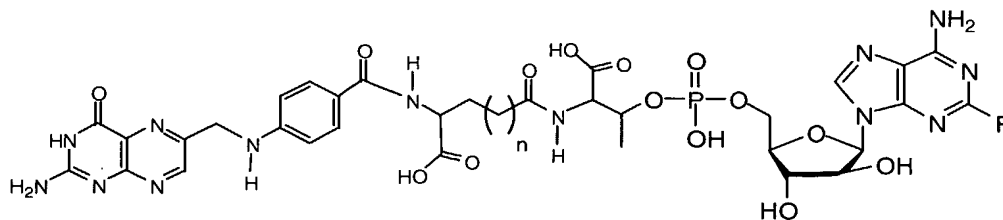
24

wherein "n" is an integer from 0 to about 200.

In another embodiment, the invention features a compound having Formula 25:

25
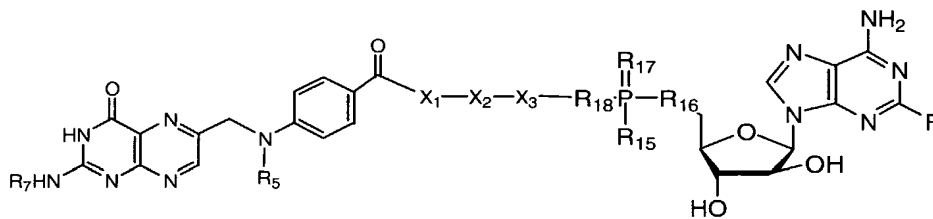
wherein each R_5 and R_7 is independently hydrogen, alkyl or a nitrogen protecting group, each R_{15}, R_{16}, R_{17}, and R_{18} is independently O, S, alkyl, substituted alkyl, aryl, substituted aryl, or halogen, X_1 is $-CH(X_{1'})$ or a group of Formula 38:

38
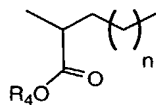
wherein R_4 is a protecting group and "n" is an integer from 0 to about 200;

$X_{1'}$ is the protected or unprotected side chain of a naturally occurring or non-naturally-occurring amino acid, X_2 is amide, alkyl, or carbonyl containing linker or a bond, and X_3 is a degradable linker which is optionally absent.

In another embodiment, the X_3 group of Formula 25 comprises a group of Formula 26:
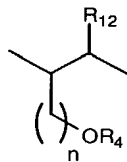
26

wherein R_4 is hydrogen or a protecting group, "n" is an integer from 0 to about 200 and R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl.

In yet another embodiment, R_4 of Formula 26 is hydrogen and R_{12} is methyl or hyrdogen.

In still another embodiment, the invention features a compound having Formula 27:

27
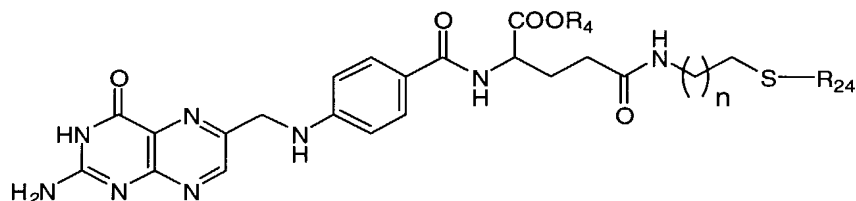
wherein "n" is an integer from about 0 to about 20, R_4 is H or a cationic salt, and R_{24} is a sulfur containing leaving group, for example a group comprising:
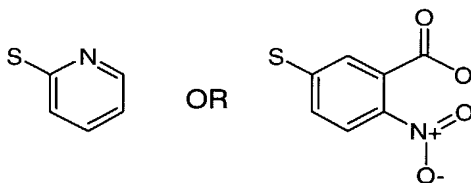
In another embodiment, the invention features a method for synthesizing a compound having Formula 27 comprising:

(a) selective tritylation of the thiol of cysteamine under conditions suitable to yield a compound having Formula 28:

28

wherein "n" is an integer from about 0 to about 20 and R_{19} is a thiol protecting group;

(b) peptide coupling of the product of (a) with a compound having Formula 29:

$$R_{21}HN \diagup \diagdown \diagup COOH$$
$$|$$
$$COOR_{20}$$

29

wherein R_{20} is a carboxylic acid protecting group and R_{21} is an amino protecting group, under conditions suitable to yield a compound having Formula 30:

$$R_{21}HN \diagup \diagup \diagup \overset{\overset{O}{\|}}{C} \diagdown \underset{H}{N} \diagdown \diagdown \overset{}{\left(\,\right)_{n}} SR_{19}$$
$$|$$
$$COOR_{20}$$

30

wherein "n" is an integer from about 0 to about 20, R_{19} is a thiol protecting group, R_{20} is a carboxylic acid protecting group and R_{21} is an amino protecting group;

(c) removing the amino protecting group R_{21} of the product of (b) under conditions suitable to yield a compound having Formula 31:

$$H_{2}N \diagup \diagup \diagup \overset{\overset{O}{\|}}{C} \diagdown \underset{H}{N} \diagdown \diagdown \overset{}{\left(\,\right)_{n}} SR_{19}$$
$$|$$
$$COOR_{20}$$

31

wherein "n" is an integer from about 0 to about 20 and R_{19} and R_{20} are as described in (b);

(d) condensation of the product of (c) with a compound having Formula 32:

$$R_{22}HN \diagbox COOH$$

32

wherein R_{22} is an amino protecting group, under conditions suitable to yield a compound having Formula 33:

33

wherein "n" is an integer from about 0 to about 20 and R_{19} and R_{20} are as described in (b) and R_{22} is as described in (d);

(e) selective cleavage of R_{22} from the product of (d) under conditions suitable to yield a compound having Formula 34:

34
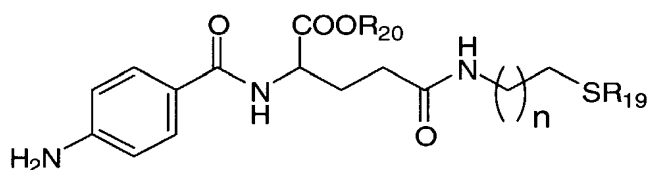
wherein "n" is an integer from about 0 to about 20 and R_{19} and R_{20} are as described in (b);

(f) coupling the product of (e) with a compound having Formula 35:

35
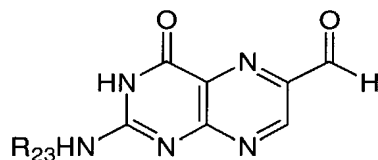
wherein R_{23} is an amino protecting group under conditions suitable to yield a compound having Formula 36:

36
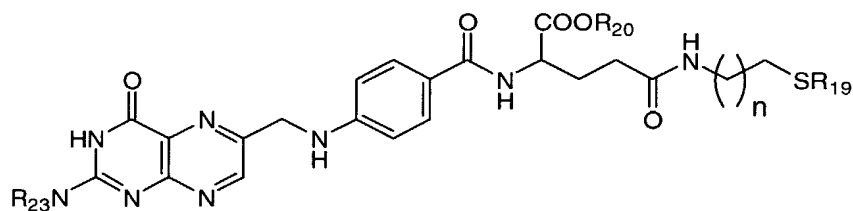
wherein R_{23} is an amino protecting group, "n" is an integer from about 0 to about 20 and R_{19} and R_{20} are as described in (b);

(g) deprotecting the product of (f) under conditions suitable to yield a compound having Formula 37.

37
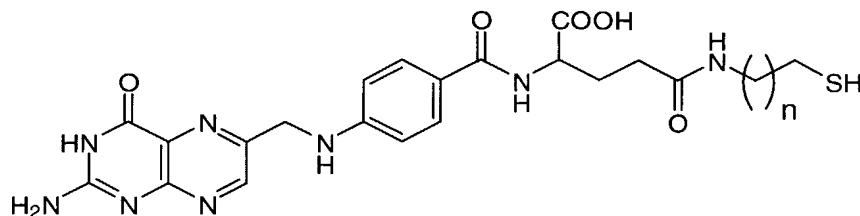
wherein "n" is an integer from about 0 to about 20; and

(h) introducing a disulphide-based leaving group to the product of (g) under conditions suitable to yield a compound having Formula 27.

In one embodiment, the invention features a compound having Formula 39:

39

wherein "n" is an integer from about 0 to about 20, X is a nucleic acid, polynucleotide, or oligonucleotide, and P is a phosphorus containing group.

In another embodiment, the invention features a method for synthesizing a compound having Formula 39, comprising:

(a) Coupling a thiol containing linker to a nucleic acid, polynucleotide or oligonucleotide under conditions suitable to yield a compound having Formula 40:
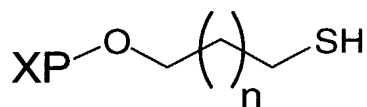
40

wherein "n" is an integer from about 0 to about 20, X is a nucleic acid, polynucleotide, or oligonucleotide, and P is a phosphorus containing group; and

(b) coupling the product of (a) with a compound having Formula 37 under conditions suitable to yield a compound having Formula 39.

5   In another embodiment, the thiol containing linker of the invention is a compound having Formula 41:
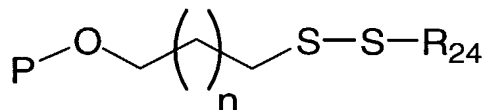
$$P{-}O{-}\underset{n}{\left(\right)}{-}S{-}S{-}R_{24}$$

41

wherein "n" is an integer from about 0 to about 20, P is a phosphorus containing
10   group, for example a phosphine, phosphite, or phosphate, and R_{24} is any alkyl, substituted alkyl, alkoxy, aryl, substituted aryl, alkenyl, substituted alkenyl, alkynyl, or substituted alkynyl group with or without additional protecting groups.

In another embodiment, the conditions suitable to yield a compound having Formula 40 comprises reduction, for example using dithiothreitol (DTT) or any
15   equivalent disulphide reducing agent, of the disulfide bond of a compound having Formula 42:
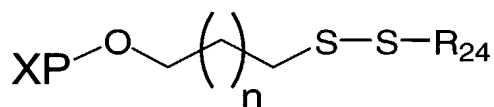
$$XP{-}O{-}\underset{n}{\left(\right)}{-}S{-}S{-}R_{24}$$

42

wherein "n" is an integer from about 0 to about 20, X is a nucleic acid,
20   polynucleotide, or oligonucleotide, P is a phosphorus containing group, and R_{24} is any alkyl, substituted alkyl, alkoxy, aryl, substituted aryl, alkenyl, substituted alkenyl, alkynyl, or substituted alkynyl group with or without additional protecting groups.

In one embodiment, the invention features a compound having Formula 43:

$$X{-}W{-}\left[Y{-}\left(CH_2CH_2O\right)_n\right]_{N'}{-}Z$$

43

wherein X comprises a biologically active molecule; W comprises a degradable nucleic acid linker; Y comprises a linker molecule or amino acid that can be present or absent; Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; n is an integer from about 1 to about 100; and N' is an integer from about 1 to about 20.

In another embodiment, the invention features a compound having Formula 44:

$$X\!-\!W\!-\!NH \overset{\displaystyle O}{\underset{\displaystyle (CH_2)_n}{\overset{\|}{\big\langle}}} \; HN \overset{O}{\overset{\|}{\diagup}} O\!-\!PEG$$

44

wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent; n is an integer from about 1 to about 50, and PEG represents a compound having Formula 45:

$$-\!\!\left[\!CH_2CH_2O\!\right]_n\!\!-\!Z$$

45

wherein Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; and n is an integer from about 1 to about 100.

In another embodiment, the invention features a compound having Formula 46:

$$X—Y—NH—CH_2CH_2—N \begin{cases} —CH_2CH_2CH_2—N—W—PEG \\ \\ —CH_2CH_2—N—W—PEG \end{cases}$$

46

wherein X comprises a biologically active molecule; each W independently comprises linker molecule or chemical linkage that can be present or absent, Y comprises

5   a linker molecule or chemical linkage that can be present or absent; and PEG represents a compound having Formula 45:

$$—\left[CH_2CH_2O\right]_n—Z$$

45
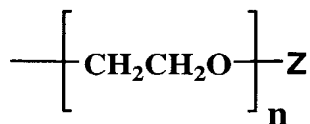
wherein Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl,

10   substituted aryl, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; and n is an integer from about 1 to about 100.

In one embodiment, the invention features a compound having Formula 47:

$$X—W—Y—R_1—\overset{\overset{R_4}{\|}}{\underset{R_2}{P}}—R_3—W—\left(Q\right)_n$$

15   47

wherein X comprises a biologically active molecule; each W independently comprises a linker molecule or chemical linkage that can be the same or different and can be present or absent, Y comprises a linker molecule that can be present or absent; each Q independently comprises a hydrophobic group or phospholipid; each R1, R2, R3, and R4

20   independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-

alkyl, S-alkylcyano, N or substituted N, and n is an integer from about 1 to about 10.

In another embodiment, the invention features a compound having Formula 48:

$$X\!-\!\!-\!W\!-\!\!-\!Y\!-\!R_1\!-\!\overset{\overset{R_4}{\|}}{\underset{\underset{R_2}{|}}{P}}\!-\!R_3 \begin{cases} R_1\!-\!\overset{\overset{R_4}{\|}}{\underset{\underset{R_2}{|}}{P}}\!-\!R_3\!-\!W\!-\!B \\[2em] R_1\!-\!\overset{\overset{R_4}{|}}{\underset{\underset{R_2}{\|}}{P}}\!-\!R_3\!-\!W\!-\!B \end{cases}$$
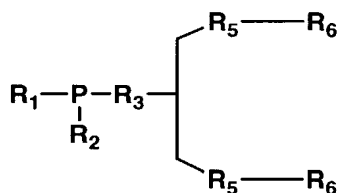
48

wherein X comprises a biologically active molecule; each W independently comprises a linker molecule or chemical linkage that can be present or absent, Y comprises a linker molecule that can be present or absent; each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, and B represents a lipophilic group, for example a saturated or unsaturated linear, branched, or cyclic alkyl group.

In another embodiment, the invention features a compound having Formula 49:

$$X\!-\!\!-\!W\!-\!\!-\!Y\!-\!R_1\!-\!\overset{\overset{R_4}{\|}}{\underset{\underset{R_2}{|}}{P}}\!-\!R_3 \overset{O\!-\!\!-\!B}{\underset{O\!-\!\!-\!B}{\diagup}}$$

49

wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent, Y comprises a linker molecule that can be present or absent; each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, and B represents a lipophilic group, for example a saturated or unsaturated linear, branched, or cyclic alkyl group..

In another embodiment, the invention features a compound having Formula 50:

$$X—Y—NH—CH_2CH_2—N(—CH_2CH_2CH_2—N—W—Q)(—CH_2CH_2—N—W—Q)$$

50

wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent, Y comprises a linker molecule or chemical linkage that can be present or absent; and each Q independently comprises a hydrophobic group or phospholipid.

In one embodiment, the invention features a compound having Formula 51:

$$X—W(Z)[—Y—SG]_n$$

51

wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent; Y comprises a linker molecule or amino acid that can be present or absent; Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; SG comprises a sugar, for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose, fructose, or fucose and the respective D or L, alpha or beta isomers, and n is an integer from about 1 to about 20.

In another embodiment, the invention features a compound having Formula 52:

52

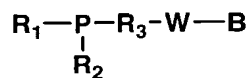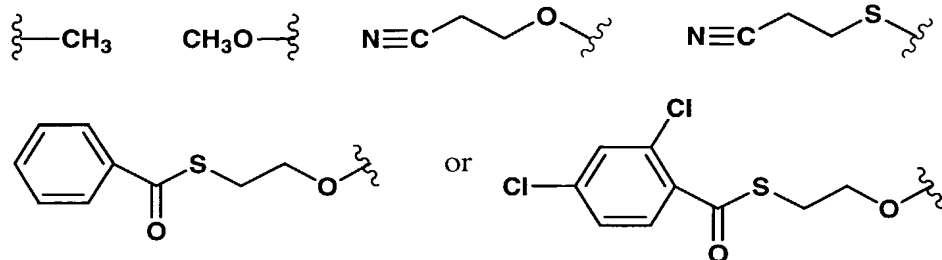wherein X comprises a biologically active molecule; Y comprises a linker molecule or chemical linkage that can be present or absent; each R1, R2, R3, R4, and R5 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N; Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; SG comprises a sugar, for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose, fructose, or fucose and the respective D or L, alpha or beta isomers, n is an integer from about 1 to about 20; and N' is an integer from about 1 to about 20..

In another embodiment, the invention features a compound having Formula 53:

53

wherein B comprises H, a nucleoside base, or a non-nucleosidic base with or without protecting groups; each R1 independently comprises O, N, S, alkyl, or substituted N; each R2 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylhalo, S, N, substituted N, or a phosphorus containing group; each R3 independently comprises N or O-N, each R4 independently comprises O, CH2, S, sulfone, or sulfoxy; X comprises H, a removable protecting group, amino, substituted amino, nucleotide, nucleoside, nucleic

acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, or label; W comprises a linker molecule or chemical linkage that can be present or absent; SG comprises a sugar, for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose, fructose, or fucose and the respective D or L, alpha or

5   beta isomers,, each n is independently an integer from about 1 to about 50; and N' is an integer from about 1 to about 10.

In another embodiment, the invention features a compound having Formula 54:

54

10   wherein B comprises H, a nucleoside base, or a non-nucleosidic base with or without protecting groups; each R1 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylhalo, S, N, substituted N, or a phosphorus containing group; X comprises H, a removable protecting group, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid,

15   phospholipid, or label; W comprises a linker molecule or chemical linkage that can be present or absent; and SG comprises a sugar, for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose, fructose, or fucose and the respective D or L, alpha or beta isomers.

In one embodiment, the invention features a compound having Formula 55:

20

55

wherein each R1 independently comprises O, N, S, alkyl, or substituted N; each R2 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylhalo, S, N, substituted N, or a phosphorus containing group; each R3 independently comprises H, OH, alkyl, substituted alkyl, or halo; X comprises H, a removable protecting group, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid, phosphol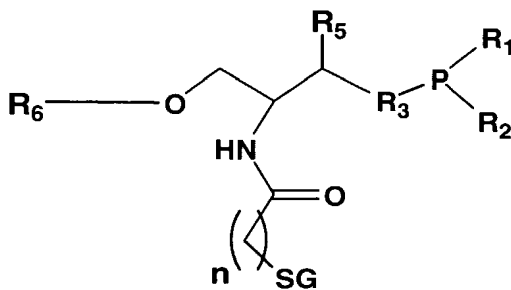ipid, biologically active molecule or label; W comprises a linker molecule or chemical linkage that can be present or absent; SG comprises a sugar, for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose, fructose, or fucose and the respective D or L, alpha or beta isomers, each n is independently an integer from about 1 to about 50; and N' is an integer from about 1 to about 100.

In another embodiment, the invention features a compound having Formula 56:

56

wherein R1 comprises H, alkyl, alkylhalo, N, substituted N, or a phosphorus containing group; R2 comprises H, O, OH, alkyl, alkylhalo, halo, S, N, substituted N, or a phosphorus containing group; X comprises H, a removable protecting group, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, biologically active molecule or label; W comprises a linker molecule or chemical linkage that can be present or absent; SG comprises a sugar, for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose, fructose, or fucose and the respective D or L, alpha or beta isomers, and each n is independently an integer from about 0 to about 20.

In another embodiment, the invention features a compound having Formula 57:

57

wherein R1 can include the groups:

or

5   and wherein R2 can include the groups:

or

and wherein Tr is a removable protecting group, for example a trityl, monomethoxytrityl, or dimethoxytrityl; SG comprises a sugar, for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose, fructose, or fucose and the
10   respective D or L, alpha or beta isomers, and n is an integer from about 1 to about 20.

In one embodiment, compounds having Formula 52, 53, 54, 55, 56, and 57 are featured wherein each nitrogen adjacent to a carbonyl can independently be substituted for a carbonyl adjacent to a nitrogen or each carbonyl adjacent to a nitrogen can be substituted for a nitrogen adjacent to a carbonyl.

15   In another embodiment, the invention features a compound having Formula 58:

$$X\text{—}W\text{—}\left[Y\text{—}\left(V\right)_n\right]N'$$

58

wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent; Y comprises a linker molecule or amino acid that can be present or absent; V comprises a protein or peptide, for example Human serum albumin protein, Antennapedia peptide, Kaposi fibroblast growth factor peptide, Caiman crocodylus Ig(5) light chain peptide, HIV envelope glycoprotein gp41 peptide, HIV-1 Tat peptide, Influenza hemagglutinin envelope glycoprotein peptide, or transportan A peptide; each n is independently an integer from about 1 to about 50; and N' is an integer from about 1 to about 100.

In another embodiment, the invention features a compound having Formula 59:

$$Q-R_2 \diagup \diagup \underset{H}{N} \overset{O}{\overset{\|}{C}} (CH_2)_n \diagup \diagdown O-N-W-\!\!-\!\!-V$$

$$\underset{\underset{X}{\overset{|}{W}}}{\overset{|}{R_1}}$$

59

wherein each R1 independently comprises O, S, N, substituted N, or a phosphorus containing group; each R2 independently comprises O, S, or N; X comprises H, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, or enzymatic nucleic acid or other biologically active molecule; n is an integer from about 1 to about 50, Q comprises H or a removable protecting group which can be optionally absent, each W independently comprises a linker molecule or chemical linkage that can be present or absent, and V comprises a protein or peptide, for example Human serum albumin protein, Antennapedia peptide, Kaposi fibroblast growth factor peptide, Caiman crocodylus Ig(5) light chain peptide, HIV envelope glycoprotein gp41 peptide, HIV-1 Tat peptide, Influenza hemagglutinin envelope glycoprotein peptide, or transportan A peptide, or a compound having Formula 45

$$-\left[CH_2CH_2O\right]_n-Z$$

45

wherein Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl, amino, substituted amino, nucleotide, nucleoside, nucleic acid,

oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; and n is an integer from about 1 to about 100..

In another embodiment, the invention features a compound having Formula 60:

5

60

wherein R1 can include the groups:

and wherein R2 can include the groups:

10   and wherein Tr is a removable protecting group, for example a trityl, monomethoxytrityl, or dimethoxytrityl; n is an integer from about 1 to about 50; and R8 is a nitrogen protecting group, for example a phthaloyl, trifluoroacetyl, FMOC, or monomethoxytrityl group.

In another embodiment, the invention features a compound having Formula 61:

$$X-W-Y-R_1-\overset{\overset{R_4}{\|}}{\underset{\underset{R_2}{|}}{P}}-R_3-W-(V)_n$$

15

61

SUBSTITUTE SHEET (RULE 26)

wherein X comprises a biologically active molecule; each W independently comprises a linker molecule or chemical linkage that can be the same or different and can be present or absent, Y comprises a linker molecule that can be present or absent; each 5 independently comprises a protein or peptide, for example Human serum albumin protein, Antennapedia peptide, Kaposi fibroblast growth factor peptide, Caiman crocodylus Ig(5) light chain peptide, HIV envelope glycoprotein gp41 peptide, HIV-1 Tat peptide, Influenza hemagglutinin envelope glycoprotein peptide, or transportan A peptide;; each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, and n is an integer from about 1 to about 10.

In another embodiment, the invention features a compound having Formula 62:

62

wherein X comprises a biologically active molecule; each 5 independently comprises a protein or peptide, for example Human serum albumin protein, Antennapedia peptide, Kaposi fibroblast growth factor peptide, Caiman crocodylus Ig(5) light chain peptide, HIV envelope glycoprotein gp41 peptide, HIV-1 Tat peptide, Influenza hemagglutinin envelope glycoprotein peptide, or transportan A peptide; W comprises a linker molecule or chemical linkage that can be present or absent; each R1, R2, and R3 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, and each n is independently an integer from about 1 to about 10.

In another embodiment, the invention features a compound having Formula 63:

63

wherein X comprises a biologically active molecule; V comprises a protein or peptide, for example Human serum albumin protein, Antennapedia peptide, Kaposi fibroblast growth factor peptide, Caiman crocodylus Ig(5) light chain peptide, HIV envelope glycoprotein gp41 peptide, HIV-1 Tat peptide, Influenza hemagglutinin envelope glycoprotein peptide, or transportan A peptide; W comprises a linker molecule or chemical linkage that can be present or absent; each R1, R2, R3 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, R4 represents an ester, amide, or protecting group, and each n is independently an integer from about 1 to about 10.

In another embodiment, the invention features a compound having Formula 64:

$$X-W-Y-R_1-\overset{\overset{R_4}{\|}}{\underset{\underset{R_2}{|}}{P}}-R_3 \begin{cases} R_1-\overset{\overset{R_4}{\|}}{\underset{\underset{R_2}{|}}{P}}-R_3-W-A \\ \\ R_1-\overset{\overset{R_4}{\|}}{\underset{\underset{R_2}{|}}{P}}-R_3-W-B \end{cases}$$

64

wherein X comprises a biologically active molecule; each W independently comprises a linker molecule or chemical linkage that can be present or absent, Y comprises a linker molecule that can be present or absent; each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, A comprises a nitrogen containing group, and B comprises a lipophilic group.

In another embodiment, the invention features a compound having Formula 65:

$$X-W-Y-R_1-\overset{\overset{R_4}{\|}}{\underset{\underset{R_2}{|}}{P}}-R_3 \begin{cases} W-R_5 \\ \\ W-R_6 \end{cases}$$

65

wherein X comprises a biologically active molecule; each W independently comprises a linker molecule or chemical linkage that can be present or absent, Y comprises a linker molecule that can be present or absent; each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, RV comprises the lipid or phospholipid component of any of Formulae 47-50, and R6 comprises a nitrogen containing group.

In another embodiment, the invention features a compound having Formula 92:

92

wherein B comprises H, a nucleoside base, or a non-nucleosidic base with or without protecting groups; each R1 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylhalo, S, N, substituted N, or a phosphorus containing group; X comprises H, a removable protecting group, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, biologically active molecule or label; W comprises 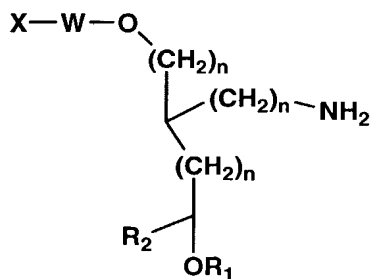a linker molecule or chemical linkage that can be present or absent; R2 comprises O, NH, S, CO, COO, ON=C, or alkyl; R3 comprises alkyl, akloxy, or an aminoacyl side chain; and SG comprises a sugar, for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose, fructose, or fucose and the respective D or L, alpha or beta isomers.

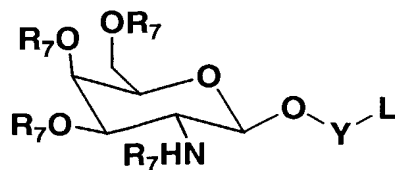In another embodiment, the invention features a compound having Formula 86:
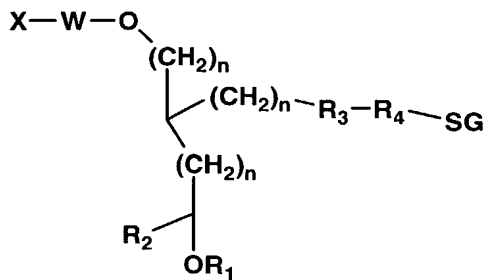
86

wherein R1 comprises H, alkyl, alkylhalo, N, substituted N, or a phosphorus containing group; R2 comprises H, O, OH, alkyl, alkylhalo, halo, S, N, substituted N, or a phosphorus containing group; X comprises H, a removable protecting group, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, biologically active molecule or label; W comprises a linker molecule or chemical linkage that can be present or absent; R3 comprises O, NH, S, CO, COO, ON=C, or alkyl; R4 comprises alkyl, akloxy, or an aminoacyl side chain; and SG comprises a sugar, for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose, fructose, or fucose and the respective D or L, alpha or beta isomers, and each n is independently an integer from about 0 to about 20.

In another embodiment, the invention features a compound having Formula 87:

$$Y \text{——} W \text{——} \underset{\underset{R_1}{|}}{C} \equiv N \text{——} O \text{——} X$$

87

wherein X comprises a protein, peptide, antibody, lipid, phospholipid, oligosaccharide, label, biologically active molecule, for example a vitamin such as folate, vitamin A, E, B6, B12, coenzyme, antibiotic, antiviral, nucleic acid, nucleotide, nucleoside, or oligonucleotide such as an enzymatic nucleic acid, allozyme, antisense nucleic acid, siRNA, 2,5-A chimera, decoy, aptamer or triplex forming oligonucleotide, or polymers such as polyethylene glycol; W comprises a linker molecule or chemical linkage that can be present or absent; and Y comprises a biologically active molecule, for example an enzymatic nucleic acid, allozyme, antisense nucleic acid, siRNA, 2,5-A chimera, decoy, aptamer or triplex forming oligonucleotide, peptide, protein, or antibody; R1 comprises H, alkyl, or substituted alkyl.

In another embodiment, the invention features a compound having Formula 88:

$$Y \text{——} W \text{——} \overset{\overset{\displaystyle O}{\|}}{C} \text{——} NH \text{——} O \text{——} X$$

88

wherein X comprises a protein, peptide, antibody, lipid, phospholipid, oligosaccharide, label, biologically active molecule, for example a vitamin such as folate, vitamin A, E, B6, B12, coenzyme, antibiotic, antiviral, nucleic acid, nucleotide,

nucleoside, or oligonucleotide such as an enzymatic nucleic acid, allozyme, antisense nucleic acid, siRNA, 2,5-A chimera, decoy, aptamer or triplex forming oligonucleotide, or polymers such as polyethylene glycol; W comprises a linker molecule or chemical linkage that can be present or absent, and Y comprises a biologically active molecule, for example an enzymatic nucleic acid, allozyme, antisense nucleic acid, siRNA, 2,5-A chimera, decoy, aptamer or triplex forming oligonucleotide, peptide, protein, or antibody.

In one embodiment, the invention features a method for the synthesis of a compound having Formula 48:

$$
X-W-Y-R_1-\overset{\overset{\textstyle R_4}{\|}}{\underset{\underset{\textstyle R_2}{|}}{P}}-R_3
\begin{cases}
R_1-\overset{\overset{\textstyle R_4}{\|}}{\underset{\underset{\textstyle R_2}{|}}{P}}-R_3-W-B \\
\\
R_1-\overset{\overset{\textstyle R_4}{|}}{\underset{\underset{\textstyle R_2}{\|}}{P}}-R_3-W-B
\end{cases}
$$

48

wherein X comprises a biologically active molecule; each W independently comprises a linker molecule or chemical linkage that can be present or absent, Y comprises a linker molecule that can be present or absent; each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N; and each B independently represents a lipophilic group, for example a saturated or unsaturated linear, branched, or cyclic alkyl group, comprising: (a) introducing a compound having Formula 66:

$$
R_1-\underset{\underset{\textstyle R_2}{|}}{P}-R_3
\begin{cases}
R_5-R_6 \\
\\
R_5-R_6
\end{cases}
$$

66

wherein R1 is defined as in Formula 48 and can include the groups:

and wherein R2 is defined as in Formula 48 and can include the groups:

and wherein each R5 independently comprises O, N, or S and each R6
5   independently comprises a removable protecting group, for example a trityl,
monomethoxytrityl, or dimethoxytrityl group, to a compound having Formula 67:

$$X \text{——} W \text{——} Y$$

67

wherein X comprises a biologically active molecule; W comprises a linker
10   molecule or chemical linkage that can be present or absent, and Y comprises a linker
molecule that can be present or absent, under conditions suitable for the formation of a
compound having Formula 68:

68

15   wherein X comprises a biologically active molecule; W comprises a linker
molecule or chemical linkage that can be present or absent, Y comprises a linker molecule
that can be present or absent; and each R1, R2, R3, and R4 independently comprises O,
OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or
substituted N comprising, each R5 independently comprises O, S, or N; and each R6 is
20   independently a removable protecting group, for example a trityl, monomethoxytrityl, or

dimethoxytrityl group; (b) removing R6 from the compound having Formula 26 and (c) introducing a compound having Formula 69:

$$R_1 - \underset{\underset{R_2}{|}}{P} - R_3 - W - B$$

69

5   wherein R1 is defined as in Formula 48 and can include the groups:

$$\text{—CH}_3 \qquad CH_3O\text{—} \qquad N{\equiv}C\diagup\diagup O\diagdown \qquad N{\equiv}C\diagup\diagup S\diagdown$$

and wherein R2 is defined as in Formula 48 and can include the groups:

$$\text{—N}\diagup \qquad \text{—N}\diagdown_{\displaystyle CH_2CH_3}^{\displaystyle CH_2CH_3} \qquad \text{—N}\diagup \qquad \text{—N}\bigcirc \qquad \text{or} \qquad \text{—N}\diagdown O$$
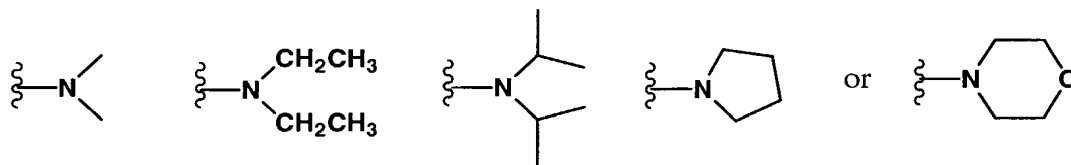
        and wherein W and B are defined as in Formula 48, to the compound having
10   Formula 68 under conditions suitable for the formation of a compound having Formula
48.

        In another embodiment, the invention features a method for the synthesis of a compound having Formula 49:

$$X - W - Y - R_1 - \underset{\underset{R_2}{|}}{\overset{\overset{R_4}{||}}{P}} - R_3 - \overset{\displaystyle R_5 - B}{\underset{\displaystyle R_5 - B}{\diagup}}$$
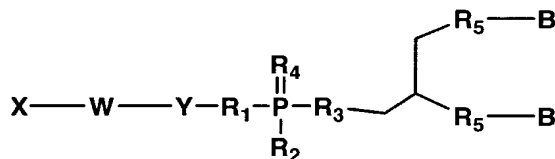
15   49

        wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent, Y comprises a linker molecule that can be present or absent; each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or
20   substituted N; each R5 independently comprises O, S, or N; and each B

independently comprises a lipophilic group, for example a saturated or unsaturated linear, branched, or cyclic alkyl group, comprising: (a) coupling a compound having Formula 70:

5   70

wherein R1 is defined as in Formula 49 and can include the groups:
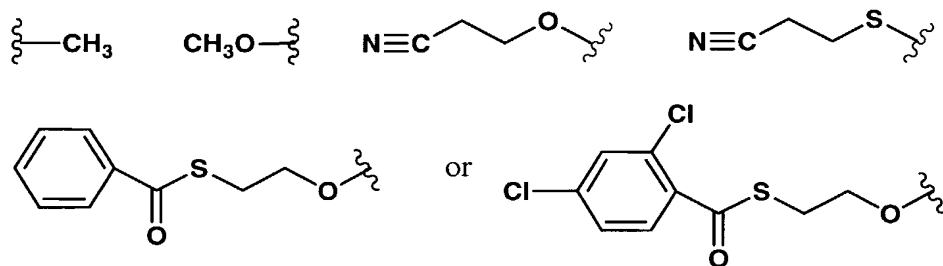
and wherein R2 is defined as in Formula 49 and can include the groups:
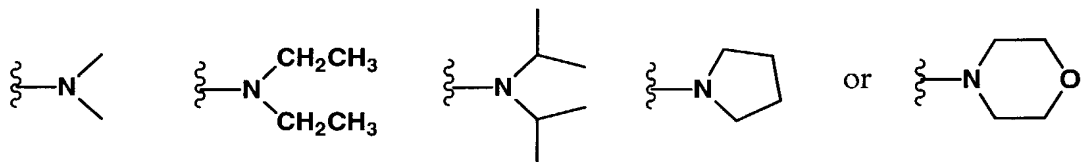
10   and wherein each R5 independently comprises O, S, or N, and wherein each B independently comprises a lipophilic group, for example a saturated or unsaturated linear, branched, or cyclic alkyl group, with a compound having Formula 67:

$$X\text{——}W\text{——}Y$$

15   67

wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent, and Y comprises a linker molecule that can be present or absent, under conditions suitable for the formation of a compound having Formula 49.

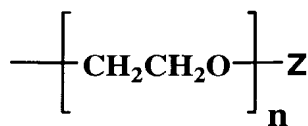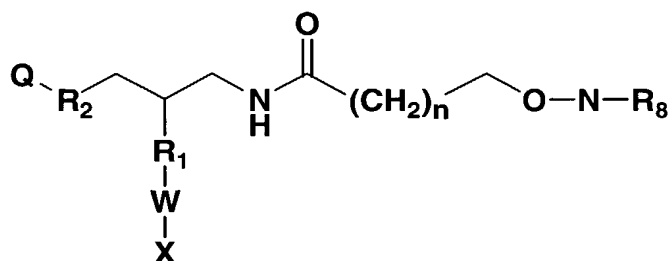In another embodiment, the invention features a method for the synthesis of a compound having Formula 52:

52

5   wherein X comprises a biologically active molecule; Y comprises a linker molecule or chemical linkage that can be present or absent; each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N; Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl, amino, substituted amino, nucleotide,
10   nucleoside, nucleic acid, oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; SG comprises a sugar, for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose, fructose, or fucose and the respective D or L, alpha or beta isomers, n is an integer from about 1 to about 20; and N' is an integer from about 1 to about 20, comprising: (a) coupling a compound having Formula 71:

71

15

wherein R1, R2, R3, R5, SG, and n is as defined in Formula 52, and wherein R1 can include the groups:

and wherein R2 can include the groups:

and R6 comprises a removable protecting group, for example a trityl,
monomethoxytrityl, or dimethoxytrityl group; with a compound having Formula 72:

$$X - Y$$

72

wherein X comprises a biologically active molecule and Y comprises a linker
molecule that can be present or absent, under conditions suitable for the formation of a
compound having Formula 95:

95

(b) removing R6 from the compound having Formula 95 and (c) optionally
coupling a nucleotide, nucleoside, nucleic acid, oligonucleotide, amino acid, peptide,
protein, lipid, phospholipid, or label, or optionally; coupling a compound having Formula
71 under and optionally repeating (b) and (c) under conditions suitable for the formation
of a compound having Formula 52.

In another embodiment, the invention features a method for synthesizing a compound having Formula 53:
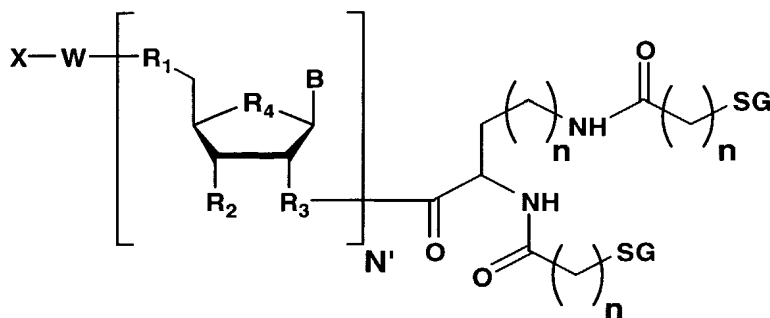
53

5   wherein B comprises H, a nucleoside base, or a non-nucleosidic base with or without protecting groups; each R1 independently comprises O, N, S, alkyl, or substituted N; each R2 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylhalo, S, N, substituted N, or a phosphorus containing group; each R3 independently comprises N or O-N, each R4 independently comprises O,

10   CH2, S, sulfone, or sulfoxy; X comprises H, a removable protecting group, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, or label; W comprises a linker molecule or chemical linkage that can be present or absent; SG comprises a sugar, for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose,

15   fructose, or fucose and the respective D or L, alpha or beta isomers, each n is independently an integer from about 1 to about 50; and N' is an integer from about 1 to about 10, comprising: coupling a compound having Formula 73:
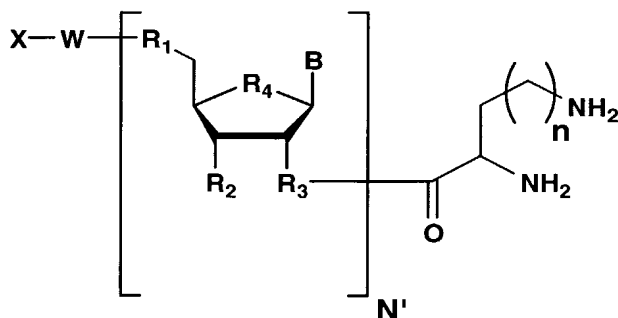
73

20   wherein R1, R2, R3, R4, X, W, B, N' and n are as defined in Formula 53, with a sugar, for example a compound having Formula 74:

74
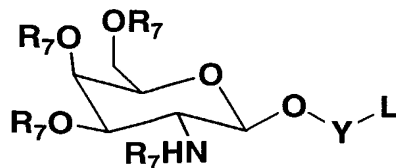
wherein Y comprises a linker molecule or chemical linkage that can be present or
absent; L represents a reactive chemical group, for example a NHS ester, and each R7
independently comprises an acyl group that can be present or absent, for example a
acetyl group; under conditions suitable for the formation of a compound having
Formula 53.

In another embodiment, the invention features a method for the synthesis of a
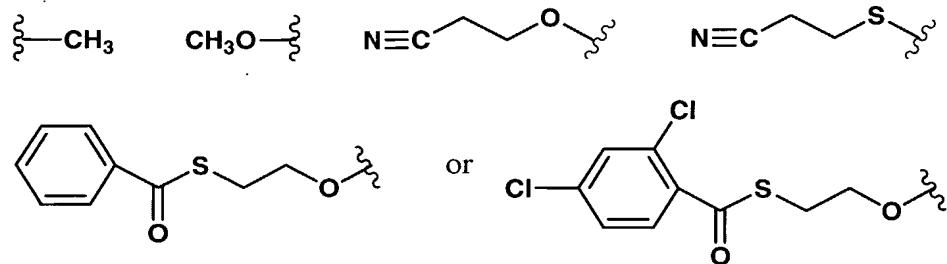compound having Formula 54:

54

wherein B comprises H, a nucleoside base, or a non-nucleosidic base with
or without protecting groups; each R1 independently comprises O, OH, H, alkyl,
alkylhalo, O-alkyl, O-alkylhalo, S, N, substituted N, or a phosphorus containing
group; X comprises H, a removable protecting group, amino, substituted amino,
nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid,
peptide, protein, lipid, phospholipid, biologically active molecule or label; W comprises a
linker molecule or chemical linkage that can be present or absent; SG comprises a sugar,
for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose,
fructose, or fucose and the respective D or L, alpha or beta isomers, comprising (a)
coupling a compound having Formula 75:
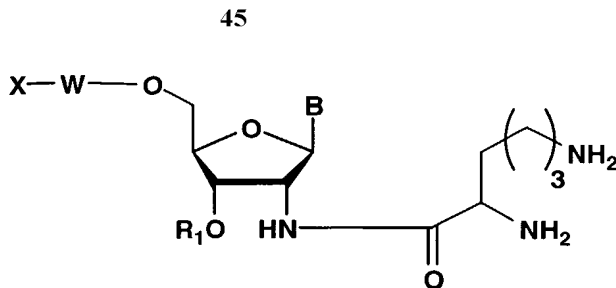
75

wherein R1, R2, R3, R4, X, W, and B are as defined in Formula 53, with a sugar, for example a compound having Formula 74 .

74
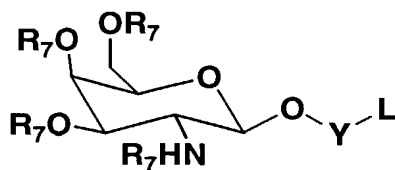
wherein Y comprises a C11 alkyl linker molecule; L represents a reactive chemical group, for example a NHS ester, and each R7 independently comprises an acyl group that can be present or absent, for example a acetyl group; under conditions suitable for the formation of a compound having Formula 54.

In another embodiment, the invention features a method for the synthesis of a compound having Formula 55:

55

wherein each R1 independently comprises O, N, S, alkyl, or substituted N; each R2 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylhalo, S, N, substituted N, or a phosphorus containing group; each R3 independently comprises H, OH, alkyl, substituted alkyl, or halo; X comprises H, a

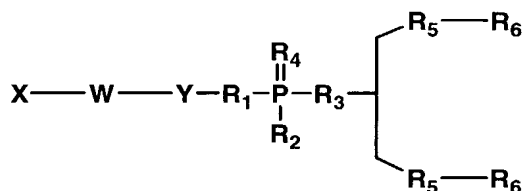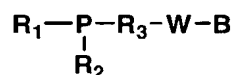removable protecting group, nucleotide, nucleoside, nucleic acid, oligonucleotide, or enzymatic nucleic acid or biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent; SG comprises a sugar, for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose, fructose, or fucose

5 and the respective D or L, alpha or beta isomers, each n is independently an integer from about 1 to about 50; and N' is an integer from about 1 to about 100, comprising: (a) coupling a compound having Formula 76:

76

10 wherein R1 can include the groups:

and wherein R2 can include the groups:

and wherein each R3 independently comprises H, OH, alkyl, substituted

15 alkyl, or halo; SG comprises a sugar, for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose, fructose, or fucose and the respective D or L, alpha or beta isomers, and n is an integer from about 1 to about 20, to a compound X-W, wherein X comprises a nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, biologically active molecule or

20 label, and W comprises a linker molecule or chemical linkage that can be present or

absent; and (b) optionally repeating step (a) under conditions suitable for the formation of a compound having Formula 55.

In another embodiment, the invention features a method for the synthesis of a compound having Formula 56:

56

wherein R1 comprises H, alkyl, alkylhalo, N, substituted N, or a phosphorus containing group; R2 comprises H, O, OH, alkyl, alkylhalo, halo, S, N, substituted N, or a phosphorus containing group; X comprises H, a removable protecting group, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, biologically active molecule or label; W comprises a linker molecule or chemical linkage that can be present or absent; SG comprises a sugar, for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose, fructose, or fucose and the respective D or L, alpha or beta isomers, and each n is independently an integer from about 0 to about 20, comprising: (a) coupling a compound having Formula 77:

77

wherein each R1, X, W, and n are as defined in Formula 56, to a sugar, for example a compound having Formula 74:

48

74

wherein Y comprises an alkyl linker molecule of length n, where n is an integer from about 1 to about 20; L represents a reactive chemical group, for example a NHS ester, and each R7 independently comprises an acyl group that can be present or absent, for example a acetyl group; and (b) optionally coupling X-W, wherein X comprises a removable protecting group, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, or label and W comprises a linker molecule or chemical linkage that can be present or absent, under conditions suitable for the formation of a compound having Formula 56.

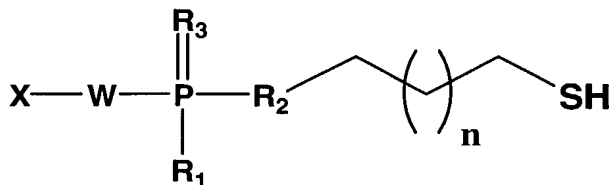In another embodiment, the invention features method for synthesizing a compound having Formula 57:
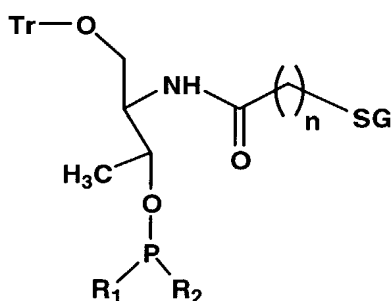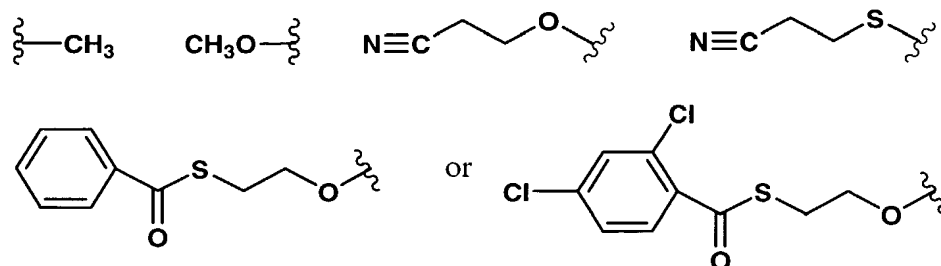
57

wherein R1 can include the groups:

and wherein R2 can include the groups:

and wherein Tr is a removable protecting group, for example a trityl, monomethoxytrityl, or dimethoxytrityl; SG comprises a sugar, for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose, fructose, or fucose and the respective D or L, alpha or beta isomers, and n is an integer from about 1 to about 20, comprising: (a) coupling a compound having Formula 77:

77
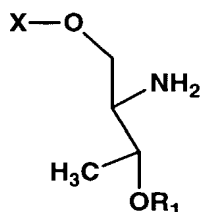
wherein R1 and X comprise H, to a sugar, for example a compound having Formula 74:
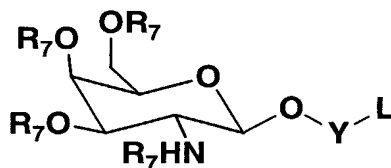
74

wherein Y comprises an alkyl linker molecule of length n, where n is an integer from about 1 to about 20; L represents a reactive chemical group, for example a NHS ester, and each R7 independently comprises an acyl group that can be present or absent, for example a acetyl group; and (b) introducing a trityl group, for example a dimethoxytrityl, monomethoxytrityl, or trityl group to the primary hydroxyl of the product of (a) and (c) introducing a phosphorus containing group having Formula 78:

78

wherein R1 can include the groups:
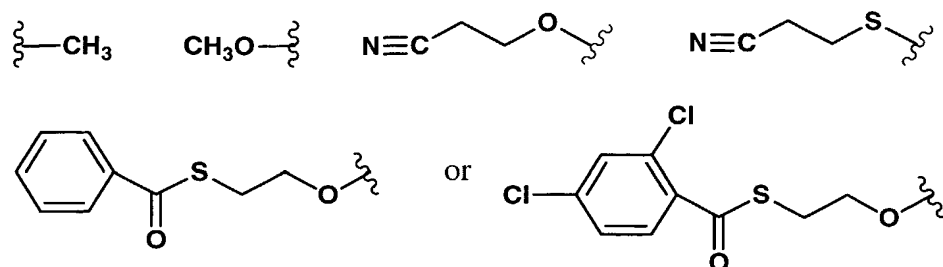
and wherein each R2 and R3 independently can include the groups:
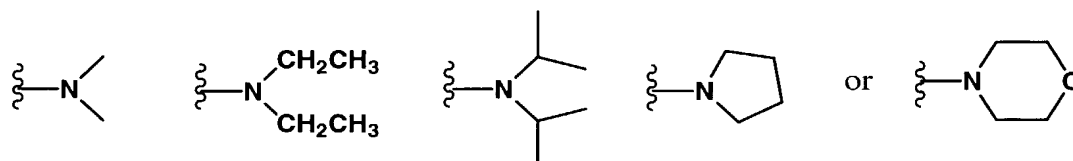
5   to the secondary hydroxyl of the product of (b) under conditions suitable for the formation of a compound having Formula 57.

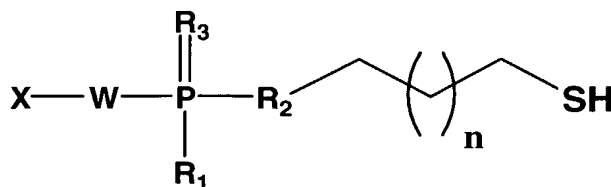In another embodiment, the invention features a method for synthesizing a compound having Formula 60:
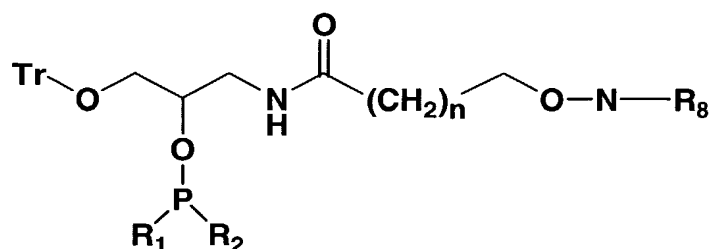
10   60

wherein R1 can include the groups:
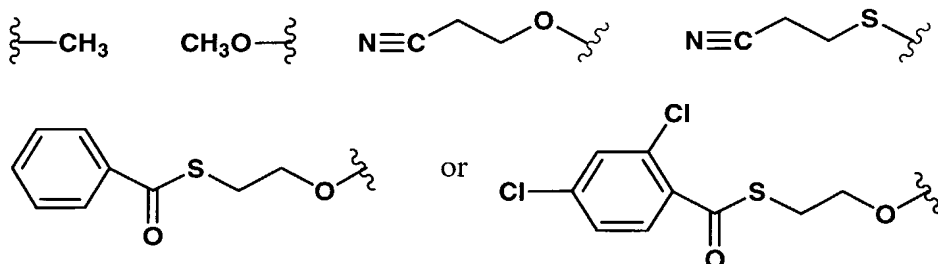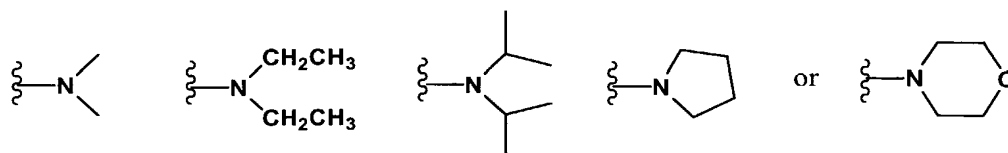
and wherein R2 can include the groups:

and wherein Tr is a removable protecting group, for example a trityl, monomethoxytrityl, or dimethoxytrityl; n is an integer from about 1 to about 50; and R8 is a nitrogen protecting group, for example a phthaloyl, trifluoroacetyl, FMOC, or monomethoxytrityl group, comprising: (a) introducing carboxy protection to a compound having Formula 79:
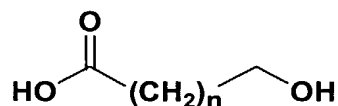
79 .
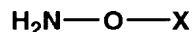
wherein n is an integer from about 1 to about 50, under conditions suitable for the formation of a compound having Formula 80:

80
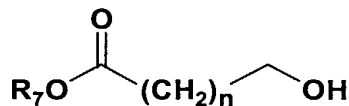
wherein n is an integer from about 1 to about 50 and R7 is a carboxylic acid protecting group, for example a benzyl group; (b) introducing a nitrogen containing group to the product of (a) under conditions suitable for the formation of a compound having Formula 81:
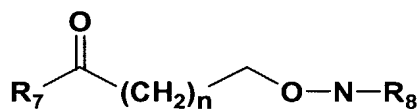
82

wherein n and R7 are as defined in Formula 80 and R8 is a nitrogen protecting group, for example a phthaloyl, trifluoroacetyl, FMOC, or monomethoxytrityl group; (c) removing the carboxylic acid protecting group from the product of (b) and introducing aminopropanediol under conditions suitable for the formation of a compound having Formula 82:

$$\text{HO} \diagup \underset{\underset{\text{HO}}{|}}{\diagdown} \diagup \text{NH} \diagdown \underset{\overset{\text{O}}{\|}}{\text{C}} \diagdown (\text{CH}_2)_n \diagdown \diagup \text{O} - \text{N} - \text{R}_8$$

82
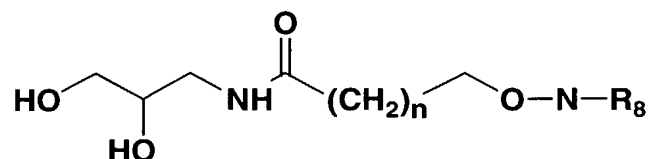
wherein n and R8 are as defined in Formula 81; (d) introducing a removable protecting group, for example a trityl, monomethoxytrityl, or dimethoxytrityl to the product of (c) under conditions suitable for the formation of a compound having Formula 83:

$$\text{TrO} \diagup \underset{\underset{\text{HO}}{|}}{\diagdown} \diagup \text{NH} \diagdown \underset{\overset{\text{O}}{\|}}{\text{C}} \diagdown (\text{CH}_2)_n \diagdown \diagup \text{O} - \text{N} - \text{R}_8$$

83

wherein Tr, n and R8 are as defined in Formula 60; and (e) introducing a phosphorus containing group having Formula 78:

$$\underset{R_1}{\overset{\displaystyle R_3}{\underset{\diagup \quad \diagdown}{\underset{\displaystyle \quad}{\overset{\displaystyle |}{P}}}}} \quad R_2$$

78

wherein R1 can include the groups:

$$\text{---CH}_3 \qquad \text{CH}_3\text{O---} \qquad \text{N}{\equiv}\text{C} \diagup \diagdown \diagup \text{O} \diagdown \qquad \text{N}{\equiv}\text{C} \diagup \diagdown \diagup \text{S} \diagdown$$

[benzoyl-S-CH2CH2-O structure]   or   [dichlorobenzoyl-S-CH2CH2-O structure]

and wherein each R2 and R3 independently can include the groups:

$$\text{---N}{\diagup}^{\diagdown} \qquad \text{---N}{\diagdown}^{\text{CH}_2\text{CH}_3}_{\text{CH}_2\text{CH}_3} \qquad \text{---N}{\diagdown}^{\diagup} \qquad \text{---N}\langle \text{pyrrolidine} \rangle \quad \text{or} \quad \text{---N}\langle \text{morpholine} \rangle$$

to the product of (d) under conditions suitable for the formation of a compound having Formula 60.

In another embodiment, the invention features a method for the synthesis of a compound having Formula 59:

$$Q{\sim}R_2\text{---}\underset{\underset{\underset{X}{\overset{|}{W}}}{\overset{|}{R_1}}}{\overset{}{}}\text{---}\underset{H}{N}\text{---}\overset{O}{\overset{\|}{C}}\text{---}(CH_2)_n\text{---}O\text{---}N\text{---}W\text{------}V$$

59

wherein each R1 independently comprises O, S, N, substituted N, or a phosphorus containing group; each R2 independently comprises O, S, or N; X comprises H, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid or biologically active molecule; n is an integer from about 1 to about 50, Q comprises H or a removable protecting group which can be optionally absent, each W independently comprises a linker molecule or chemical linkage that can be present or absent, and V comprises a protein or peptide, for example Human serum albumin protein, Antennapedia peptide, Kaposi fibroblast growth factor peptide, Caiman crocodylus Ig(5) light chain peptide, HIV envelope glycoprotein gp41 peptide, HIV-1 Tat peptide, Influenza hemagglutinin envelope glycoprotein peptide, or transportan A peptide, or a compound having Formula 45:

$$\text{---}\left[CH_2CH_2O\right]_n\text{---}Z$$

45

wherein Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; and n is an integer from about 1 to about 100, comprising: (a) removing R8 from a compound having Formula 84:

$$ Q{\sim}R_2 \frown \overset{\displaystyle R_1}{\underset{\displaystyle \underset{X}{\overset{|}{W}}}{|}} \frown \overset{H}{\underset{N}{}} \overset{O}{\overset{\|}{C}} (CH_2)_n \frown O{-}N{-}R_8 $$

84

wherein Q, X, W, R1, R2, and n are as defined in Formula 59 and R8 is a nitrogen protecting group, for example a phthaloyl, trifluoroacetyl, FMOC, or monomethoxytrityl group, under conditions suitable for the formation of a compound having Formula 85:

$$ Q{\sim}R_2 \frown \overset{\displaystyle R_1}{\underset{\displaystyle \underset{X}{\overset{|}{W}}}{|}} \frown \overset{H}{\underset{N}{}} \overset{O}{\overset{\|}{C}} (CH_2)_n \frown O{-}NH_2 $$

85

wherein Q, X, W, R1, R2, and n are as defined in Formula 59; (b) introducing a group V to the product of (a) via the formation of an oxime linkage, wherein V comprises a protein or peptide, for example Human serum albumin protein, Antennapedia peptide, Kaposi fibroblast growth factor peptide, Caiman crocodylus Ig(5) light chain peptide, HIV envelope glycoprotein gp41 peptide, HIV-1 Tat peptide, Influenza hemagglutinin envelope glycoprotein peptide, or transportan A peptide, or a compound having Formula 45:

$$ {-}{\Big[}CH_2CH_2O{\Big]}_n{-}Z $$

45

wherein Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; and n is an integer from about 1 to about 100, under conditions suitable for the formation of a compound having Formula 59.

In another embodiment, the invention features a method for synthesizing a compound having Formula 64:

$$
X\!-\!\!-\!W\!-\!\!-\!Y\!-\!R_1\!-\!\overset{\overset{\displaystyle R_4}{\|}}{\underset{\underset{\displaystyle R_2}{|}}{P}}\!-\!R_3
\begin{cases}
R_1\!-\!\overset{\overset{\displaystyle R_4}{\|}}{\underset{\underset{\displaystyle R_2}{|}}{P}}\!-\!R_3\!-\!W\!-\!A \\[2em]
R_1\!-\!\overset{\overset{\displaystyle R_4}{\|}}{\underset{\underset{\displaystyle R_2}{|}}{P}}\!-\!R_3\!-\!W\!-\!B
\end{cases}
$$

64

wherein X comprises a biologically active molecule; each W independently comprises a linker molecule or chemical linkage that can be present or absent, Y comprises a linker molecule that can be present or absent; each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, A comprises a nitrogen containing group, and B comprises a lipophilic group, comprising: (a) introducing a compound having Formula 66:

$$
R_1\!-\!\underset{\underset{\displaystyle R_2}{|}}{P}\!-\!R_3
\begin{cases}
R_5\!-\!\!-\!R_6 \\[1.5em]
R_5\!-\!\!-\!R_6
\end{cases}
$$

66

wherein R1 is defined as in Formula 64 and can include the groups:

$$\xi\!-\!CH_3 \qquad CH_3O\!-\!\xi \qquad N\!\equiv\!C\!\smile\!\!O\!\smile \qquad N\!\equiv\!C\!\smile\!\!S\!\smile$$

[structure: phenyl-C(=O)-S-CH2CH2-O—]   or   [structure: 2,4-dichlorophenyl-C(=O)-S-CH2CH2-O—]

and wherein R2 is defined as in Formula 64 and can include the groups:

$$\overset{\xi}{\xi}-N\diagup\diagdown \qquad \overset{\xi}{\xi}-N\overset{CH_2CH_3}{\underset{CH_2CH_3}{\diagdown}} \qquad \overset{\xi}{\xi}-N\diagup\diagdown \qquad \overset{\xi}{\xi}-N\diagup\diagdown \qquad or \qquad \overset{\xi}{\xi}-N\diagup\diagdown O$$

and wherein each R5 independently comprises O, N, or S and each R6 independently comprises a removable protecting group, for example a trityl, monomethoxytrityl, or dimethoxytrityl group, to a compound having Formula 67:

5

$$X-\!\!\!-\!\!\!-W-\!\!\!-\!\!\!-Y$$

67

wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent, and Y comprises a linker molecule that can be present or absent, under conditions suitable for the formation of a

10   compound having Formula 68:

$$X-\!\!\!-W-\!\!\!-Y-R_1-\overset{\overset{R_4}{\|}}{\underset{R_2}{P}}-R_3-\Big[\begin{matrix} R_5-R_6 \\ \\ R_5-R_6 \end{matrix}$$

68

wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent, Y comprises a linker molecule

15   that can be present or absent; and each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N comprising, each R5 independently comprises O, S, or N; and each R6 is independently a removable protecting group, for example a trityl, monomethoxytrityl, or dimethoxytrityl group; (b) removing R6 from the compound having Formula 68 and (c)

20   introducing a compound having Formula 69:

$$R_1-\overset{}{\underset{R_2}{P}}-R_3-W-B$$

69

wherein R1 is defined as in Formula 64 and can include the groups:

and wherein R2 is defined as in Formula 64 and can include the groups:

and wherein R3, W and B are defined as in Formula 64; and introducing a
5   compound having Formula 69':

$$R_1-P-R_3-W-A$$
$$\underset{R_2}{|}$$

69'
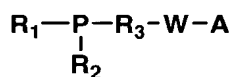
wherein R1 is defined as in Formula 64 and can include the groups:
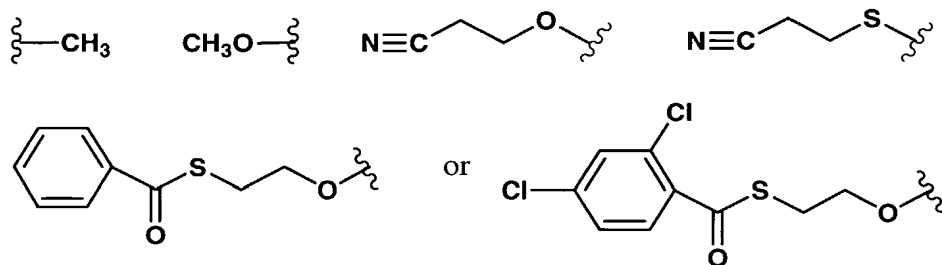
10   and wherein R2 is defined as in Formula 48 and can include the groups:
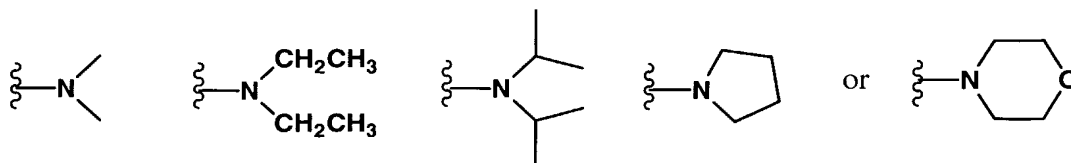
and wherein R3, W and A are defined as in Formula 64; to the compound having
Formula 68 under conditions suitable for the formation of a compound having Formula
64.

In another embodiment, the invention features a method for the synthesis of a compound having Formula 62:
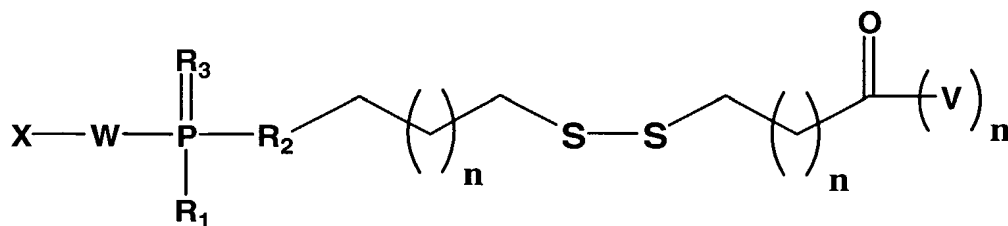
62

5   wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent; each S independently comprises a protein or peptide, for example Human serum albumin protein, Antennapedia peptide, Kaposi fibroblast growth factor peptide, Caiman crocodylus Ig(5) light chain peptide, HIV envelope glycoprotein gp41 peptide, HIV-1 Tat peptide, Influenza

10   hemagglutinin envelope glycoprotein peptide, or transportan A peptide;; each R1, R2, and R3 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, and each n is independently an integer from about 1 to about 10, comprising: (a) introducing a compound having Formula 93:
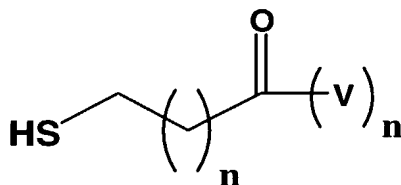
93

15

wherein V and n are as defined in Formula 62, to a compound having Formula 86:

86

wherein X, W, R1, R2, R3, and n are as defined in Formula 62, under conditions

20   suitable for the formation of a compound having Formula 62.

In another embodiment, the invention features a method for the synthesis of a compound having Formula 63:

63

5   wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent; V comprises a protein or peptide, for example Human serum albumin protein, Antennapedia peptide, Kaposi fibroblast growth factor peptide, Caiman crocodylus Ig(5) light chain peptide, HIV envelope glycoprotein gp41 peptide, HIV-1 Tat peptide, Influenza hemagglutinin
10   envelope glycoprotein peptide, or transportan A peptide;; each R1, R2, and R3 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, R4 represents an ester, amide, or protecting group, and each n is independently an integer from about 1 to about 10, comprising: (a) introducing a compound having Formula 96:

96

       wherein V and R4 are as defined in Formula 63, to a compound having Formula 86:

86

60

wherein X, W, R1, R2, R3, and n are as defined in Formula 63, under conditions suitable for the formation of a compound having Formula 63.

In another embodiment, the invention features a method for the synthesis of a compound having Formula 87:

$$Y—W—\underset{\underset{R_1}{|}}{C}{=\!=\!=}N—O—X$$

5

87

wherein X comprises a protein, peptide, antibody, lipid, phospholipid, oligosaccharide, label, biologically active molecule, for example a vitamin such as folate, vitamin A, E, B6, B12, coenzyme, antibiotic, antiviral, nucleic acid, nucleotide,
10   nucleoside, or oligonucleotide such as an enzymatic nucleic acid, allozyme, antisense nucleic acid, siRNA, 2,5-A chimera, decoy, aptamer or triplex forming oligonucleotide, or polymers such as polyethylene glycol; W comprises a linker molecule or chemical linkage that can be present or absent; and Y comprises a biologically active molecule, for example an enzymatic nucleic acid, allozyme, antisense nucleic acid, siRNA, 2,5-A chimera,
15   decoy, aptamer or triplex forming oligonucleotide, peptide, protein, or antibody; R1 comprises H, alkyl, or substituted alkyl, comprising (a) coupling a compound having Formula 89:

$$Y—W—\underset{\underset{R_1}{|}}{C}{=\!=}O$$

89

20   wherein Y, W and R are as defined in Formula 87, with a compound having Formula 90:

$$H_2N—O—X$$

90

wherein X is as defined in Formula 87, under conditions suitable for the formation
25   of a compound having Formula 87, for example by post-synthetic conjugation of a compound having Formula 89 with a compound having Formula 90, wherein X of compound 90 comprises an enzymatic nucleic acid molecule and Y of Formula 89 comprises a peptide.

In another embodiment, the invention features a method for the synthesis of a compound having Formula 88:

$$Y-W-\overset{\overset{\displaystyle O}{\|}}{C}-NH-O-X$$

88

wherein X comprises a protein, peptide, antibody, lipid, phospholipid, oligosaccharide, label, biologically active molecule, for example a vitamin such as folate, vitamin A, E, B6, B12, coenzyme, antibiotic, antiviral, nucleic acid, nucleotide, nucleoside, or oligonucleotide such as an enzymatic nucleic acid, allozyme, antisense nucleic acid, siRNA, 2,5-A chimera, decoy, aptamer or triplex forming oligonucleotide, or polymers such as polyethylene glycol; W comprises a linker molecule or chemical linkage that can be present or absent, and Y comprises a biologically active molecule, for example an enzymatic nucleic acid, allozyme, antisense nucleic acid, siRNA, 2,5-A chimera, decoy, aptamer or triplex forming oligonucleotide, peptide, protein, or antibody, comprising (a) coupling a compound having Formula 91:

$$Y-W-\underset{\underset{\displaystyle H}{|}}{C}=O$$
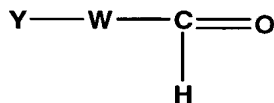
91

wherein Y and W are as defined in Formula 88, with a compound having Formula 90:

$$H_2N-O-X$$

90

wherein X is as defined in Formula 88, under conditions suitable for the formation of a compound having Formula 88, for example by post-synthetic conjugation of a compound having Formula 91 with a compound having Formula 90, wherein X of compound 90 comprises an enzymatic nucleic acid molecule and Y of Formula 91 comprises a peptide.

In one embodiment, the invention features a compound having Formula 94,

$$X-Y-W-Y-Z$$

94

wherein X comprises a protein, peptide, antibody, lipid, phospholipid, oligosaccharide, label, biologically active molecule, for example a vitamin such as folate, vitamin A, E, B6, B12, coenzyme, antibiotic, antiviral, nucleic acid, nucleotide, nucleoside, or oligonucleotide such as an enzymatic nucleic acid, allozyme, antisense nucleic acid, siRNA, 2,5-A chimera, decoy, aptamer or triplex forming oligonucleotide, or polymers such as polyethylene glycol; each Y independently comprises a linker or chemical linkage that can be present or absent, W comprises a biodegradable nucleic acid linker molecule, and Z comprises a biologically active molecule, for example an enzymatic nucleic acid, allozyme, antisense nucleic acid, siRNA, 2,5-A chimera, decoy, aptamer or triplex forming oligonucleotide, peptide, protein, or antibody.

In another embodiment, W of a compound having Formula 94 of the invention comprises 5'-cytidine-deoxythymidine-3', 5'-deoxythymidine-cytidine-3', 5'-cytidine-deoxyuridine-3', 5'-deoxyuridine-cytidine-3', 5'-uridine-deoxythymidine-3', or 5'-deoxythymidine-uridine-3'.

In yet another embodiment, W of a compound having Formula 94 of the invention comprises 5'-adenosine-deoxythymidine-3', 5'-deoxythymidine-adenosine-3', 5'-adenosine-deoxyuridine-3', or 5'-deoxyuridine-adenosine-3'.

In another embodiment, Y of a compound having Formula 94 of the invention comprises a phosphorus containing linkage, phoshoramidate linkage, phosphodiester linkage, phosphorothioate linkage, amide linkage, ester linkage, carbamate linkage, disulfide linkage, oxime linkage, or morpholino linkage.

In another embodiment, compounds having Formula 89 and 91 of the invention are synthesized by periodate oxidation of an N-terminal Serine or Threonine residue of a peptide or protein.

In one embodiment, X of compounds having Formulae 43, 44, 46-52, 58, 61-65, 85-88, 92, 94, and 95 of the invention comprises an enzymatic nucleic acid.

In another embodiment, X of compounds having Formulae 43, 44, 46-52, 58, 61-65, 85-88, 92, 94, and 95 of the invention comprises an antibody. In yet another embodiment, X of compounds having Formulae 43, 44, 46-52, 58, 61-65, 85-88, 92, 94, and 95 of the invention comprises an interferon.

In another embodiment, X of compounds having Formulae 43, 44, 46-52, 58, 61-65, 85-88, 92, 94, and 95 of the invention comprises an antisense nucleic acid, dsRNA, ssRNA, decoy, triplex oligonucleotide, aptamer, or 2,5-A chimera.

In one embodiment, W and/or Y of compounds having Formulae 43, 44, 46-56, 58-59, 61-65, 67, 68, 69, 72, 73, 75, 77, 84-89, 91-92, 94, and 95 of the invention comprises a degradable or cleavable linker, for example a nucleic acid sequence comprising ribonucleotides and/or deoxynucleotides, such as a dimer, trimer, or tetramer. A non limiting example of a nucleic acid cleavable linker is an adenosine-deoxythymidine (A-dT) dimer or a cytidine-deoxythymidine (C-dT) dimer. In yet another embodiment, W and/or V of compounds having Formulae 43, 44, 48-51, 58, 63-65, and 96 of the invention comprises a N-hydroxy succinimide (NHS) ester linkage, oxime linkage, disulfide linkage, phosphoramidate, phosphorothioate, phosphorodithioate, phosphodiester linkage, or $NHC(O)$, $CH_3NC(O)$, $CONH$, $C(O)NCH_3$, S, SO, SO_2, O, NH, NCH_3 group. In another embodiment, the degradable linker, W and/or Y, of compounds having Formulae 43, 44, 46-56, 58-59, 61-65, 67, 68, 69, 72, 73, 75, 77, 84-89, 91-92, 94, and 95 of the invention comprises a linker that is susceptible to cleavage by carboxypeptidase activity.

In another embodiment, W and/or Y of Formulae 43, 44, 46-56, 58-59, 61-65, 67, 68, 69, 72, 73, 75, 77, 84-89, 91-92, 94, and 95 comprises a polyethylene glycol linker having Formula 45:
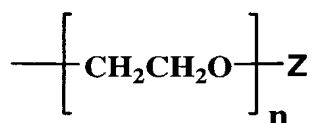
$$-\left[CH_2CH_2O\right]_n-Z$$

45

wherein Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; and n is an integer from about 1 to about 100.

In one embodiment, the nucleic acid conjugates of the instant invention are assembled by solid phase synthesis, for example on an automated peptide synthesizer, for example a Miligen 9050 synthesizer and/or an automated oligonucleotide synthesizer such as an ABI 394, 390Z, or Pharmacia OligoProcess, OligoPilot, OligoMax, or AKTA synthesizer. In another embodiment, the nucleic acid conjugates of the invention are

assembled post synthetically, for example, following solid phase oligonucleotide synthesis (see for example **Figure 15**).

In another embodiment, V of compounds having Formula 58-63 and 96 comprise peptides having SEQ ID NOS: 14-23 (**Table 3**).

5   In one embodiment, the nucleic acid conjugates of the instant invention are assembled post synthetically, for example, following solid phase oligonucleotide synthesis.

The present invention provides compositions and conjugates comprising nucleosidic and non-nucleosidic derivatives. The present invention also provides nucleic
10   acid, polynucleotide and oligonucleotide derivatives including RNA, DNA, and PNA based conjugates. The attachment of compounds of the invention to nucleosides, nucleotides, non-nucleosides, and nucleic acid molecules is provided at any position within the molecule, for example, at internucleotide linkages, nucleosidic sugar hydroxyl groups such as 5', 3', and 2'-hydroxyls, and/or at nucleobase positions such as amino and
15   carbonyl groups.

The exemplary conjugates of the invention are described as compounds of the formulae herein, however, other peptide, protein, phospholipid, and poly-alkyl glycol derivatives are provided by the invention, including various analogs of the compounds of formulae 1-96, including but not limited to different isomers of the compounds described
20   herein.

In one embodiment, the present invention features molecules, compositions and conjugates of molecules, for example, non-nucleosidic small molecules, nucleosides, nucleotides, and nucleic acids, such as enzymatic nucleic acid molecules, antisense nucleic acids, 2-5A antisense chimeras, triplex oligonucleotides, decoys, siRNA,
25   allozymes, aptamers, and antisense nucleic acids containing RNA cleaving chemical groups.

The exemplary folate conjugates of the invention are described as compounds shown by formulae herein, however, other folate and antifolate derivatives are provided by the invention, including various folate analogs of the formulae of the invention,
30   including   dihydrofloates,   tetrahydrofolates,   tetrahydorpterins,   folinic   acid, pteropolyglutamic acid, 1-deza, 3-deaza, 5-deaza, 8-deaza, 10-deaza, 1,5-deaza, 5,10 dideaza, 8,10-dideaza, and 5,8-dideaza folates, antifolates, and pteroic acids. As used

herein, the term "folate" is meant to refer to folate and folate derivatives, including pteroic acid derivatives and analogs.

The present invention features compositions and conjugates to facilitate delivery of molecules into a biological system such as cells. The conjugates provided by the instant invention can impart therapeutic activity by transferring therapeutic compounds across cellular membranes. The present invention encompasses the design and synthesis of novel agents for the delivery of molecules, including but not limited to small molecules, lipids, nucleosides, nucleotides, nucleic acids, negatively charged polymers and other polymers, for example proteins, peptides, carbohydrates, or polyamines. In general, the transporters described are designed to be used either individually or as part of a multi-component system. The compounds of the invention generally shown in Formulae herein are expected to improve delivery of molecules into a number of cell types originating from different tissues, in the presence or absence of serum.

In another embodiment, the present invention features methods to modulate gene expression, for example, genes involved in the progression and/or maintenance of cancer or in a viral infection. For example, in one embodiment, the invention features the use of one or more of the nucleic acid-based molecules and methods independently or in combination to inhibit the expression of the gene(s) encoding proteins associated with cancerous conditions, for example breast cancer, lung cancer, colorectal cancer, brain cancer, esophageal cancer, stomach cancer, bladder cancer, pancreatic cancer, cervical cancer, head and neck cancer, ovarian cancer, melanoma, lymphoma, glioma, or multidrug resistant cancer associated genes.

In another embodiment, the invention features the use of one or more of the nucleic acid-based molecules and methods independently or in combination to inhibit the expression of the gene(s) encoding viral proteins, for example HIV, HBV, HCV, CMV, RSV, HSV, poliovirus, influenza, rhinovirus, west nile virus, Ebola virus, foot and mouth virus, and papilloma virus associated genes.

In one embodiment, the invention features the use of an enzymatic nucleic acid molecule conjugate comprising compounds of formulae 1-96, preferably in the hammerhead, NCH, G-cleaver, amberzyme, zinzyme and/or DNAzyme motif, to inhibit the expression of cancer and virus associated genes.

In another embodiment, the invention features the use of an enzymatic nucleic acid molecule as a conjugate. These enzymatic nucleic acids can catalyze the hydrolysis of RNA phosphodiester bonds in trans (and thus can cleave other RNA molecules) under

physiological conditions.   **Table I** summarizes some of the characteristics of these enzymatic nucleic acids.   Without being bound by any particular theory, in general, enzymatic nucleic acids act by first binding to a target RNA.   Such binding occurs through the target binding portion of an enzymatic nucleic acid which is held in close

5   proximity to an enzymatic portion of the molecule that acts to cleave the target RNA. Thus, the enzymatic nucleic acid first recognizes and then binds a target RNA through complementary base-pairing, and once bound to the correct site, acts enzymatically to cut the target RNA.   Strategic cleavage of such a target RNA destroys its ability to direct synthesis of an encoded protein.   After an enzymatic nucleic acid has bound and cleaved

10   its RNA target, it is released from that RNA to search for another target and can repeatedly bind and cleave new targets. Thus, a single enzymatic nucleic acid molecule is able to cleave many molecules of target RNA.   In addition, the enzymatic nucleic acid is a highly specific inhibitor of gene expression, with the specificity of inhibition depending not only on the base-pairing mechanism of binding to the target RNA, but also on the

15   mechanism of target RNA cleavage.   Single mismatches, or base-substitutions, near the site of cleavage can completely eliminate catalytic activity of an enzymatic nucleic acid.

In one embodiment of the invention described herein, the enzymatic nucleic acid molecule component of the conjugate is formed in a hammerhead or hairpin motif, but can also be formed in the motif of a hepatitis delta virus, group I intron, group II intron or

20   RNase P RNA (in association with an RNA guide sequence), Neurospora VS RNA, DNAzymes, NCH cleaving motifs, or G-cleavers.   Examples of such hammerhead motifs are described by Dreyfus, supra, Rossi et al., 1992, AIDS Research and Human Retroviruses 8, 183; of hairpin motifs by Hampel et al., EP0360257, Hampel and Tritz, 1989 Biochemistry 28, 4929, Feldstein et al., 1989, Gene 82, 53, Haseloff and Gerlach,

25   1989, Gene, 82, 43, and Hampel et al., 1990 Nucleic Acids Res. 18, 299; Chowrira & McSwiggen, US. Patent No. 5,631,359; of the hepatitis delta virus motif is described by Perrotta and Been, 1992 Biochemistry 31, 16; of the RNase P motif by Guerrier-Takada et al., 1983 Cell 35, 849; Forster and Altman, 1990, Science 249, 783; Li and Altman, 1996, Nucleic Acids Res. 24, 835; Neurospora VS RNA ribozyme motif is described by Collins

30   (Saville and Collins, 1990 Cell 61, 685-696; Saville and Collins, 1991 Proc. Natl. Acad. Sci. USA 88, 8826-8830; Collins and Olive, 1993 Biochemistry 32, 2795-2799; Guo and Collins, 1995, EMBO. J. 14, 363); Group II introns are described by Griffin et al., 1995, Chem. Biol. 2, 761; Michels and Pyle, 1995, Biochemistry 34, 2965; Pyle et al., International PCT Publication No. WO 96/22689; of the Group I intron by Cech et al.,

35   U.S. Patent 4,987,071 and of DNAzymes by Usman et al., International PCT Publication No. WO 95/11304; Chartrand et al., 1995, NAR 23, 4092; Breaker et al., 1995, Chem. Bio. 2, 655; Santoro et al., 1997, PNAS 94, 4262, and Beigelman et al., International PCT

publication No. WO 99/55857. NCH cleaving motifs are described in Ludwig & Sproat, International PCT Publication No. WO 98/58058; and G-cleavers are described in Kore et al., 1998, Nucleic Acids Research 26, 4116-4120 and Eckstein et al., International PCT Publication No. WO 99/16871. Additional motifs such as the Aptazyme (Breaker et al.,

5   WO 98/43993), Amberzyme (Class I motif; Figure 3; Beigelman et al., U.S. Serial No. 09/301,511) and Zinzyme (Figure 4) (Beigelman et al., U.S. Serial No. 09/301,511), all incorporated by reference herein including drawings, can also be used in the present invention. These specific motifs are not limiting in the invention and those skilled in the art will recognize that all that is important in an enzymatic nucleic acid molecule of this

10   invention is that it has a specific substrate binding site which is complementary to one or more of the target gene RNA regions, and that it have nucleotide sequences within or surrounding that substrate binding site which impart an RNA cleaving activity to the molecule (Cech et al., U.S. Patent No. 4,987,071).

In one embodiment of the present invention, a nucleic acid molecule component of

15   a conjugate of the instant invention can be between 12 and 100 nucleotides in length. For example, enzymatic nucleic acid molecules of the invention are preferably between 15 and 50 nucleotides in length, more preferably between 25 and 40 nucleotides in length, e.g., 34, 36, or 38 nucleotides in length (for example see Jarvis et al., 1996, J. Biol. Chem., 271, 29107-29112). Exemplary DNAzymes of the invention are preferably

20   between 15 and 40 nucleotides in length, more preferably between 25 and 35 nucleotides in length, e.g., 29, 30, 31, or 32 nucleotides in length (see for example Santoro et al., 1998, Biochemistry, 37, 13330-13342; Chartrand et al., 1995, Nucleic Acids Research, 23, 4092-4096). Exemplary antisense molecules of the invention are preferably between 15 and 75 nucleotides in length, more preferably between 20 and 35 nucleotides in length,

25   e.g., 25, 26, 27, or 28 nucleotides in length (see, for example, Woolf et al., 1992, PNAS., 89, 7305-7309; Milner et al., 1997, Nature Biotechnology, 15, 537-541). Exemplary triplex forming oligonucleotide molecules of the invention are preferably between 10 and 40 nucleotides in length, more preferably between 12 and 25 nucleotides in length, e.g., 18, 19, 20, or 21 nucleotides in length (see for example Maher et al., 1990, Biochemistry,

30   29, 8820-8826; Strobel and Dervan, 1990, Science, 249, 73-75). Those skilled in the art will recognize that all that is required is for the nucleic acid molecule to be of sufficient length and suitable conformation for the nucleic acid molecule to catalyze a reaction contemplated herein. The length of the nucleic acid molecules described and exemplified herein are not limiting within the general size ranges stated.

35   The conjugates of the invention are added directly, or can be complexed with cationic lipids, packaged within liposomes, or otherwise delivered to target cells or

tissues. The conjugates and/or conjugate complexes can be locally administered to relevant tissues ex vivo, or in vivo through injection or infusion pump, with or without their incorporation in biopolymers. The compositions and conjugates of the instant invention, individually, or in combination or in conjunction with other drugs, can be used

5   to treat diseases or conditions discussed above. For example, to treat a disease or condition associated with the levels of a pathogenic protein, the patient can be treated, or other appropriate cells can be treated, as is evident to those skilled in the art, individually or in combination with one or more drugs under conditions suitable for the treatment.

In a further embodiment, the described molecules can be used in combination with

10   other known treatments to treat conditions or diseases discussed above. For example, the described molecules can be used in combination with one or more known therapeutic agents to treat breast, lung, prostate, colorectal, brain, esophageal, bladder, pancreatic, cervical, head and neck, and ovarian cancer, melanoma, lymphoma, glioma, multidrug resistant cancers, and/or HIV, HBV, HCV, CMV, RSV, HSV, poliovirus, influenza,

15   rhinovirus, west nile virus, Ebola virus, foot and mouth virus, and papilloma virus infection.

Included in another embodiment are a series of multi-domain cellular transport vehicles (MCTV) including one or more compounds of Formulae 1-96 herein that enhance the cellular uptake and transmembrane permeability of negatively charged

20   molecules in a variety of cell types. The compounds of the invention are used either alone or in combination with other compounds with a neutral or a negative charge including but not limited to neutral lipid and/or targeting components, to improve the effectiveness of the formulation or conjugate in delivering and targeting the predetermined compound or molecule to cells. Another embodiment of the invention encompasses the utility of these

25   compounds for increasing the transport of other impermeable and/or lipophilic compounds into cells. Targeting components include ligands for cell surface receptors including, peptides and proteins, glycolipids, lipids, carbohydrates, and their synthetic variants, for example folate receptors.

In another embodiment, the compounds of the invention are provided as a surface

30   component of a lipid aggregate, such as a liposome encapsulated with the predetermined molecule to be delivered. Liposomes, which can be unilamellar or multilamellar, can introduce encapsulated material into a cell by different mechanisms. For example, the liposome can directly introduce its encapsulated material into the cell cytoplasm by fusing with the cell membrane. Alternatively, the liposome can be compartmentalized into an

acidic vacuole (i.e., an endosome) and its contents released from the liposome and out of the acidic vacuole into the cellular cytoplasm.

In one embodiment the invention features a lipid aggregate formulation of the compounds described herein, including phosphatidylcholine (of varying chain length; e.g., egg yolk phosphatidylcholine), cholesterol, a cationic lipid, and 1,2-distearoyl-sn-glycero-3-phosphoethanolamine-polythyleneglycol-2000 (DSPE-PEG2000). The cationic lipid component of this lipid aggregate can be any cationic lipid known in the art such as dioleoyl 1,2,-diacyl-3-trimethylammonium-propane (DOTAP). In another embodiment this cationic lipid aggregate comprises a covalently bound compound described in any of the Formulae herein.

In another embodiment, polyethylene glycol (PEG) is covalently attached to the compounds of the present invention. The attached PEG can be any molecular weight but is preferably between 2000-50,000 daltons.

The compounds and methods of the present invention are useful for introducing nucleotides, nucleosides, nucleic acid molecules, lipids, peptides, proteins, and/or non-nucleosidic small molecules into a cell. For example, the invention can be used for nucleotide, nucleoside, nucleic acid, lipids, peptides, proteins, and/or non-nucleosidic small molecule delivery where the corresponding target site of action exists intracellularly.

In one embodiment, the compounds of the instant invention provide conjugates of molecules that can interact with cellular receptors, such as high affinity folate receptors and ASGPr receptors, and provide a number of features that allow the efficient delivery and subsequent release of conjugated compounds across biological membranes. The compounds utilize chemical linkages between the receptor ligand and the compound to be delivered of length that can interact preferentially with cellular receptors. Furthermore, the chemical linkages between the ligand and the compound to be delivered can be designed as degradable linkages, for example by utilizing a phosphate linkage that is proximal to a nucleophile, such as a hydroxyl group. Deprotonation of the hydroxyl group or an equivalent group, as a result of pH or interaction with a nuclease, can result in nucleophilic attack of the phosphate resulting in a cyclic phosphate intermediate that can be hydrolyzed. This cleavage mechanism is analogous RNA cleavage in the presence of a base or RNA nuclease. Alternately, other degradable linkages can be selected that respond to various factors such as UV irradiation, cellular nucleases, pH, temperature etc. The use of degradable linkages allows the delivered compound to be released in a

predetermined system, for example in the cytoplasm of a cell, or in a particular cellular organelle.

The present invention also provides ligand derived phosphoramidites that are readily conjugated to compounds and molecules of interest. Phosphoramidite compounds
5   of the invention permit the direct attachment of conjugates to molecules of interest without the need for using nucleic acid phosphoramidite species as scaffolds. As such, the used of phosphoramidite chemistry can be used directly in coupling the compounds of the invention to a compound of interest, without the need for other condensation reactions, such as condensation of the ligand to an amino group on the nucleic acid, for
10   example at the N6 position of adenosine or a 2'-deoxy-2'-amino function. Additionally, compounds of the invention can be used to introduce non-nucleic acid based conjugated linkages into oligonucleotides that can provide more efficient coupling during oligonucleotide synthesis than the use of nucleic acid-based phosphoramidites. This improved coupling can take into account improved steric considerations of abasic or non-
15   nucleosidic scaffolds bearing pendant alkyl linkages.

Compounds of the invention utilizing triphosphate groups can be utilized in the enzymatic incorporation of conjugate molecules into oligonucleotides. Such enzymatic incorporation is useful when conjugates are used in post-synthetic enzymatic conjugation or selection reactions, (see for example Matulic-Adamic *et al.*, 2000, *Bioorg. Med. Chem.*
20   *Lett.*, 10, 1299-1302; Lee *et al.*, 2001, *NAR.*, 29, 1565-1573; Joyce, 1989, *Gene*, 82, 83-87; Beaudry *et al.*, 1992, *Science* 257, 635-641; Joyce, 1992, *Scientific American* 267, 90-97; Breaker *et al.*, 1994, *TIBTECH* 12, 268; Bartel *et al.*,1993, *Science* 261:1411-1418; Szostak, 1993, *TIBS* 17, 89-93; Kumar *et al.*, 1995, *FASEB J.*, 9, 1183; Breaker, 1996, *Curr. Op. Biotech.*, 7, 442; Santoro *et al.*, 1997, *Proc. Natl. Acad. Sci.*, 94, 4262;
25   Tang *et al.*, 1997, *RNA* 3, 914; Nakamaye & Eckstein, 1994, *supra*; Long & Uhlenbeck, 1994, supra; Ishizaka et al., 1995, *supra*; Vaish *et al.*, 1997, *Biochemistry* 36, 6495; Kuwabara *et al.*, 2000, *Curr. Opin. Chem. Biol.*, 4, 669).

Compounds of the invention can be used to detect the presence of a target molecule in a biological system, such as tissue, cell or cell lysate. Examples of target
30   molecules include nucleic acids, proteins, peptides, antibodies, polysaccharides, lipids, hormones, sugars, metals, microbial or cellular metabolites, analytes, pharmaceuticals, and other organic and inorganic molecules or other biomolecules in a sample. The compounds of the instant invention can be conjugated to a predetermined compound or molecule that is capable of interacting with the target molecule in the system and
35   providing a detectable signal or response. Various compounds and molecules known in

the art that can be used in these applications include but are not limited to antibodies, labeled antibodies, allozymes, aptamers, labeled nucleic acid probes, molecular beacons, fluorescent molecules, radioisotopes, polysaccharides, and any other compound capable of interacting with the target molecule and generating a detectable signal upon target

5   interaction.   For example, such compounds are described in Application entitled "NUCLEIC ACID SENSOR MOLECULES", USSN 09/800,594 filed on March 6, 2001 (Not yet assigned; Attorney Docket No. MBHB00-816-A 700.001) with inventors Nassim Usman and James A. McSwiggen, which is incorporated by reference in its entirety, including the drawings.

10   The term "biodegradable nucleic acid linker molecule" as used herein, refers to a nucleic acid molecule that is designed as a biodegradable linker to connect one molecule to another molecule, for example, a biologically active molecule. The stability of the biodegradable nucleic acid linker molecule can be modulated by using various combinations of ribonucleotides, deoxyribonucleotides, and chemically modified

15   nucleotides, for example 2'-O-methyl, 2'-fluoro, 2'-amino, 2'-O-amino, 2'-C-allyl, 2'-O-allyl, and other 2'-modified or base modified nucleotides. The biodegradable nucleic acid linker molecule can be a dimer, trimer, tetramer or longer nucleic acid molecule, for example an oligonucleotide of about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides in length, or can comprise a single nucleotide with a phosphorus

20   based linkage, for example a phosphoramidate or phosphodiester linkage.   The biodegradable nucleic acid linker molecule can also comprise nucleic acid backbone, nucleic acid sugar, or nucleic acid base modifications.

      The term "biodegradable" as used herein, refers to degradation in a biological system, for example enzymatic degradation or chemical degradation.

25   The term "biologically active molecule" as used herein, refers to compounds or molecules that are capable of eliciting or modifying a biological response in a system. Non-limiting examples of biologically active molecules contemplated by the instant invention include therapeutically active molecules such as antibodies, hormones, antivirals, peptides, proteins, chemotherapeutics, small molecules, vitamins, co-factors,

30   nucleosides, nucleotides, oligonucleotides, enzymatic nucleic acids, antisense nucleic acids, triplex forming oligonucleotides, 2,5-A chimeras, siRNA, dsRNA, allozymes, aptamers, decoys and analogs thereof. Biologically active molecules of the invention also include molecules capable of modulating the pharmacokinetics and/or pharmacodynamics of other biologically active molecules, for example lipids and polymers such as

35   polyamines, polyamides, polyethylene glycol and other polyethers.

The term "phospholipid" as used herein, refers to a hydrophobic molecule comprising at least one phosphorus group. For example, a phospholipid can comprise a phosphorus containing group and saturated or unsaturated alkyl group, optionally substituted with OH, COOH, oxo, amine, or substituted or unsubstituted aryl groups.

5   The term "nitrogen containing group" as used herein refers to any chemical group or moiety comprising a nitrogen or substituted nitrogen. Non-limiting examples of nitrogen containing groups include amines, substituted amines, amides, alkylamines, amino acids such as arginine or lysine, polyamines such as spermine or spermidine, cyclic amines such as pyridines, pyrimidines including uracil, thymine, and cytosine, 10   morpholines, phthalimides, and heterocyclic amines such as purines, including guanine and adenine.

The term "target molecule" as used herein, refers to nucleic acid molecules, proteins, peptides, antibodies, polysaccharides, lipids, sugars, metals, microbial or cellular metabolites, analytes, pharmaceuticals, and other organic and inorganic molecules that are 15   present in a system.

By "inhibit" or "down-regulate" it is meant that the expression of the gene, or level of RNAs or equivalent RNAs encoding one or more protein subunits, or activity of one or more protein subunits, such as pathogenic protein, viral protein or cancer related protein subunit(s), is reduced below that observed in the absence of the compounds or 20   combination of compounds of the invention. In one embodiment, inhibition or down-regulation with an enzymatic nucleic acid molecule preferably is below that level observed in the presence of an enzymatically inactive or attenuated molecule that is able to bind to the same site on the target RNA, but is unable to cleave that RNA. In another embodiment, inhibition or down-regulation with antisense oligonucleotides is preferably 25   below that level observed in the presence of, for example, an oligonucleotide with scrambled sequence or with mismatches. In another embodiment, inhibition or down-regulation of viral or oncogenic RNA, protein, or protein subunits with a compound of the instant invention is greater in the presence of the compound than in its absence.

By "up-regulate" is meant that the expression of the gene, or level of RNAs or 30   equivalent RNAs encoding one or more protein subunits, or activity of one or more protein subunits, such as viral or oncogenic protein subunit(s), is greater than that observed in the absence of the compounds or combination of compounds of the invention. For example, the expression of a gene, such as a viral or cancer related gene, can be increased in order to treat, prevent, ameliorate, or modulate a pathological condition 35   caused or exacerbated by an absence or low level of gene expression.

By "modulate" is meant that the expression of the gene, or level of RNAs or equivalent RNAs encoding one or more protein subunits, or activity of one or more protein subunit(s) of a protein, for example a viral or cancer related protein is up-regulated or down-regulated, such that the expression, level, or activity is greater than or less than that observed in the absence of the compounds or combination of compounds of the invention.

The term "enzymatic nucleic acid molecule" as used herein refers to a nucleic acid molecule which has complementarity in a substrate binding region to a specified gene target, and also has an enzymatic activity which is active to specifically cleave target RNA. That is, the enzymatic nucleic acid molecule is able to intermolecularly cleave RNA and thereby inactivate a target RNA molecule. These complementary regions allow sufficient hybridization of the enzymatic nucleic acid molecule to the target RNA and thus permit cleavage. One hundred percent complementarity is preferred, but complementarity as low as 50-75% can also be useful in this invention (see for example Werner and Uhlenbeck, 1995, Nucleic Acids Research, 23, 2092-2096; Hammann et al., 1999, Antisense and Nucleic Acid Drug Dev., 9, 25-31). The nucleic acids can be modified at the base, sugar, and/or phosphate groups. The term enzymatic nucleic acid is used interchangeably with phrases such as ribozymes, catalytic RNA, enzymatic RNA, catalytic DNA, aptazyme or aptamer-binding ribozyme, regulatable ribozyme, catalytic oligonucleotides, nucleozyme, DNAzyme, RNA enzyme, endoribonuclease, endonuclease, minizyme, leadzyme, oligozyme or DNA enzyme. All of these terminologies describe nucleic acid molecules with enzymatic activity. The specific enzymatic nucleic acid molecules described in the instant application are not limiting in the invention and those skilled in the art will recognize that all that is important in an enzymatic nucleic acid molecule of this invention is that it has a specific substrate binding site which is complementary to one or more of the target nucleic acid regions, and that it have nucleotide sequences within or surrounding that substrate binding site which impart a nucleic acid cleaving and/or ligation activity to the molecule (Cech et al., U.S. Patent No. 4,987,071; Cech et al., 1988, 260 JAMA 3030).

The term "nucleic acid molecule" as used herein, refers to a molecule having nucleotides. The nucleic acid can be single, double, or multiple stranded and can comprise modified or unmodified nucleotides or non-nucleotides or various mixtures and combinations thereof.

The term "enzymatic portion" or "catalytic domain" as used herein refers to that portion/region of the enzymatic nucleic acid molecule essential for cleavage of a nucleic acid substrate (for example see Figure 1).

The term "substrate binding arm" or "substrate binding domain" as used herein refers to that portion/region of a enzymatic nucleic acid which is able to interact, for example via complementarity (i.e., able to base-pair with), with a portion of its substrate. Preferably, such complementarity is 100%, but can be less if desired. For example, as few as 10 bases out of 14 can be base-paired (see for example Werner and Uhlenbeck, 1995, Nucleic Acids Research, 23, 2092-2096; Hammann et al., 1999, Antisense and Nucleic Acid Drug Dev., 9, 25-31). Examples of such arms are shown generally in Figures 1-4. That is, these arms contain sequences within a enzymatic nucleic acid which are intended to bring enzymatic nucleic acid and target RNA together through complementary base-pairing interactions. The enzymatic nucleic acid of the invention can have binding arms that are contiguous or non-contiguous and can be of varying lengths. The length of the binding arm(s) are preferably greater than or equal to four nucleotides and of sufficient length to stably interact with the target RNA; preferably 12-100 nucleotides; more preferably 14-24 nucleotides long (see for example Werner and Uhlenbeck, supra; Hamman et al., supra; Hampel et al., EP0360257; Berzal-Herrance et al., 1993, EMBO J., 12, 2567-73). If two binding arms are chosen, the design is such that the length of the binding arms are symmetrical (i.e., each of the binding arms is of the same length; e.g., five and five nucleotides, or six and six nucleotides, or seven and seven nucleotides long) or asymmetrical (i.e., the binding arms are of different length; e.g., six and three nucleotides; three and six nucleotides long; four and five nucleotides long; four and six nucleotides long; four and seven nucleotides long; and the like).

The term "Inozyme" or "NCH" motif as used herein, refers to an enzymatic nucleic acid molecule comprising a motif as is generally described as NCH Rz in **Figure 1**. Inozymes possess endonuclease activity to cleave RNA substrates having a cleavage triplet NCH/, where N is a nucleotide, C is cytidine and H is adenosine, uridine or cytidine, and / represents the cleavage site. H is used interchangeably with X. Inozymes can also possess endonuclease activity to cleave RNA substrates having a cleavage triplet NCN/, where N is a nucleotide, C is cytidine, and / represents the cleavage site. "I" in Figure 2 represents an Inosine nucleotide, preferably a ribo-Inosine or xylo-Inosine nucleoside.

The term "G-cleaver" motif as used herein, refers to an enzymatic nucleic acid molecule comprising a motif as is generally described as G-cleaver Rz in **Figure 1**. G-

cleavers possess endonuclease activity to cleave RNA substrates having a cleavage triplet NYN/, where N is a nucleotide, Y is uridine or cytidine and / represents the cleavage site. G-cleavers can be chemically modified as is generally shown in Figure 2.

The term "amberzyme" motif as used herein, refers to an enzymatic nucleic acid molecule comprising a motif as is generally described in **Figure 2**. Amberzymes possess endonuclease activity to cleave RNA substrates having a cleavage triplet NG/N, where N is a nucleotide, G is guanosine, and / represents the cleavage site. Amberzymes can be chemically modified to increase nuclease stability through substitutions as are generally shown in Figure 3. In addition, differing nucleoside and/or non-nucleoside linkers can be used to substitute the 5'-gaaa-3' loops shown in the figure. Amberzymes represent a non-limiting example of an enzymatic nucleic acid molecule that does not require a ribonucleotide (2'-OH) group within its own nucleic acid sequence for activity.

The term "zinzyme" motif as used herein, refers to an enzymatic nucleic acid molecule comprising a motif as is generally described in **Figure 3**. Zinzymes possess endonuclease activity to cleave RNA substrates having a cleavage triplet including but not limited to YG/Y, where Y is uridine or cytidine, and G is guanosine and / represents the cleavage site. Zinzymes can be chemically modified to increase nuclease stability through substitutions as are generally shown in **Figure 3**, including substituting 2'-O-methyl guanosine nucleotides for guanosine nucleotides. In addition, differing nucleotide and/or non-nucleotide linkers can be used to substitute the 5'-gaaa-2' loop shown in the figure. Zinzymes represent a non-limiting example of an enzymatic nucleic acid molecule that does not require a ribonucleotide (2'-OH) group within its own nucleic acid sequence for activity.

The term 'DNAzyme' as used herein, refers to an enzymatic nucleic acid molecule that does not require the presence of a 2'-OH group for its activity. In particular embodiments the enzymatic nucleic acid molecule can have an attached linker(s) or other attached or associated groups, moieties, or chains containing one or more nucleotides with 2'-OH groups. DNAzymes can be synthesized chemically or expressed endogenously in vivo, by means of a single stranded DNA vector or equivalent thereof. An example of a DNAzyme is shown in **Figure 4** and is generally reviewed in Usman et al., International PCT Publication No. WO 95/11304; Chartrand et al., 1995, NAR 23, 4092; Breaker et al., 1995, Chem. Bio. 2, 655; Santoro et al., 1997, PNAS 94, 4262; Breaker, 1999, Nature Biotechnology, 17, 422-423; and Santoro et. al., 2000, J. Am. Chem. Soc., 122, 2433-39. Additional DNAzyme motifs can be selected for using techniques similar to those described in these references, and hence, are within the scope of the present invention.

The term "sufficient length" as used herein, refers to an oligonucleotide of length great enough to provide the intended function under the expected condition, i.e., greater than or equal to 3 nucleotides. For example, for binding arms of enzymatic nucleic acid "sufficient length" means that the binding arm sequence is long enough to provide stable binding to a target site under the expected binding conditions. Preferably, the binding arms are not so long as to prevent useful turnover of the nucleic acid molecule.

The term "stably interact" as used herein, refers to interaction of the oligonucleotides with target nucleic acid (e.g., by forming hydrogen bonds with complementary nucleotides in the target under physiological conditions) that is sufficient to the intended purpose (e.g., cleavage of target RNA by an enzyme).

The term "homology" as used herein, refers to the nucleotide sequence of two or more nucleic acid molecules is partially or completely identical.

The term "antisense nucleic acid", as used herein, refers to a non-enzymatic nucleic acid molecule that binds to target RNA by means of RNA-RNA or RNA-DNA or RNA-PNA (protein nucleic acid; Egholm et al., 1993 Nature 365, 566) interactions and alters the activity of the target RNA (for a review, see Stein and Cheng, 1993 Science 261, 1004 and Woolf et al., US patent No. 5,849,902). Typically, antisense molecules are complementary to a target sequence along a single contiguous sequence of the antisense molecule. However, in certain embodiments, an antisense molecule can bind to substrate such that the substrate molecule forms a loop, and/or an antisense molecule can bind such that the antisense molecule forms a loop. Thus, the antisense molecule can be complementary to two (or even more) non-contiguous substrate sequences or two (or even more) non-contiguous sequence portions of an antisense molecule can be complementary to a target sequence or both. For a review of current antisense strategies, see Schmajuk et al., 1999, J. Biol. Chem., 274, 21783-21789, Delihas et al., 1997, Nature, 15, 751-753, Stein et al., 1997, Antisense N. A. Drug Dev., 7, 151, Crooke, 2000, Methods Enzymol., 313, 3-45; Crooke, 1998, Biotech. Genet. Eng. Rev., 15, 121-157, Crooke, 1997, Ad. Pharmacol., 40, 1-49. In addition, antisense DNA can be used to target RNA by means of DNA-RNA interactions, thereby activating RNase H, which digests the target RNA in the duplex. The antisense oligonucleotides can comprise one or more RNAse H activating region, which is capable of activating RNAse H cleavage of a target RNA. Antisense DNA can be synthesized chemically or expressed via the use of a single stranded DNA expression vector or equivalent thereof.

The term "RNase H activating region" as used herein, refers to a region (generally greater than or equal to 4-25 nucleotides in length, preferably from 5-11 nucleotides in

length) of a nucleic acid molecule capable of binding to a target RNA to form a non-covalent complex that is recognized by cellular RNase H enzyme (see for example Arrow et al., US 5,849,902; Arrow et al., US 5,989,912). The RNase H enzyme binds to the nucleic acid molecule-target RNA complex and cleaves the target RNA sequence. The RNase H activating region comprises, for example, phosphodiester, phosphorothioate (preferably at least four of the nucleotides are phosphorothiote substitutions; more specifically, 4-11 of the nucleotides are phosphorothiote substitutions); phosphorodithioate, 5'-thiophosphate, or methylphosphonate backbone chemistry or a combination thereof. In addition to one or more backbone chemistries described above, the RNase H activating region can also comprise a variety of sugar chemistries. For example, the RNase H activating region can comprise deoxyribose, arabino, fluoroarabino or a combination thereof, nucleotide sugar chemistry. Those skilled in the art will recognize that the foregoing are non-limiting examples and that any combination of phosphate, sugar and base chemistry of a nucleic acid that supports the activity of RNase H enzyme is within the scope of the definition of the RNase H activating region and the instant invention.

The term "2-5A antisense chimera" as used herein, refers to an antisense oligonucleotide containing a 5'-phosphorylated 2'-5'-linked adenylate residue. These chimeras bind to target RNA in a sequence-specific manner and activate a cellular 2-5A-dependent ribonuclease which, in turn, cleaves the target RNA (Torrence et al., 1993 Proc. Natl. Acad. Sci. USA 90, 1300; Silverman et al., 2000, Methods Enzymol., 313, 522-533; Player and Torrence, 1998, Pharmacol. Ther., 78, 55-113).

The term "triplex forming oligonucleotides" as used herein, refers to an oligonucleotide that can bind to a double-stranded DNA in a sequence-specific manner to form a triple-strand helix. Formation of such triple helix structure has been shown to inhibit transcription of the targeted gene (Duval-Valentin et al., 1992 Proc. Natl. Acad. Sci. USA 89, 504; Fox, 2000, Curr. Med. Chem., 7, 17-37; Praseuth et. al., 2000, Biochim. Biophys. Acta, 1489, 181-206).

The term "gene" it as used herein, refers to a nucleic acid that encodes an RNA, for example, nucleic acid sequences including but not limited to structural genes encoding a polypeptide.

The term "pathogenic protein" as used herein, refers to endogenous or exogenous proteins that are associated with a disease state or condition, for example a particular cancer or viral infection.

The term "complementarity" refers to the ability of a nucleic acid to form hydrogen bond(s) with another RNA sequence by either traditional Watson-Crick or other non-traditional types. In reference to the nucleic molecules of the present invention, the binding free energy for a nucleic acid molecule with its target or complementary sequence

5   is sufficient to allow the relevant function of the nucleic acid to proceed, e.g., enzymatic nucleic acid cleavage, antisense or triple helix inhibition. Determination of binding free energies for nucleic acid molecules is well known in the art (see, e.g., Turner et al., 1987, CSH Symp. Quant. Biol. LII pp.123-133; Frier et al., 1986, Proc. Nat. Acad. Sci. USA 83:9373-9377; Turner et al., 1987, J. Am. Chem. Soc. 109:3783-3785). A percent

10 complementarity indicates the percentage of contiguous residues in a nucleic acid molecule which can form hydrogen bonds (e.g., Watson-Crick base pairing) with a second nucleic acid sequence (e.g., 5, 6, 7, 8, 9, 10 out of 10 being 50%, 60%, 70%, 80%, 90%, and 100% complementary). "Perfectly complementary" means that all the contiguous residues of a nucleic acid sequence will hydrogen bond with the same number of

15 contiguous residues in a second nucleic acid sequence.

The term "RNA" as used herein, refers to a molecule comprising at least one ribonucleotide residue. By "ribonucleotide" or "2'-OH" is meant a nucleotide with a hydroxyl group at the 2' position of a β-D-ribo-furanose moiety.

The term "decoy RNA" as used herein, refers to a RNA molecule or aptamer that is

20 designed to preferentially bind to a predetermined ligand. Such binding can result in the inhibition or activation of a target molecule. The decoy RNA or aptamer can compete with a naturally occurring binding target for the binding of a specific ligand. For example, it has been shown that over-expression of HIV trans-activation response (TAR) RNA can act as a "decoy" and efficiently binds HIV tat protein, thereby preventing it from binding

25 to TAR sequences encoded in the HIV RNA (Sullenger et al., 1990, Cell, 63, 601-608). This is but a specific example and those in the art will recognize that other embodiments can be readily generated using techniques generally known in the art, see for example Gold et al., 1995, Annu. Rev. Biochem., 64, 763; Brody and Gold, 2000, J. Biotechnol., 74, 5; Sun, 2000, Curr. Opin. Mol. Ther., 2, 100; Kusser, 2000, J. Biotechnol., 74, 27;

30 Hermann and Patel, 2000, Science, 287, 820; and Jayasena, 1999, Clinical Chemistry, 45, 1628. Similarly, a decoy RNA can be designed to bind to a receptor and block the binding of an effector molecule or a decoy RNA can be designed to bind to receptor of interest and prevent interaction with the receptor.

The term "single stranded RNA" (ssRNA) as used herein refers to a naturally

35 occurring or synthetic ribonucleic acid molecule comprising a linear single strand, for

example a ssRNA can be a messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA) etc. of a gene.

The term "single stranded DNA" (ssDNA) as used herein refers to a naturally occurring or synthetic deoxyribonucleic acid molecule comprising a linear single strand, for example, a ssDNA can be a sense or antisense gene sequence or EST (Expressed Sequence Tag).

The term "double stranded RNA" or "dsRNA" as used herein refers to a double stranded RNA molecule capable of RNA interference, including short interfering RNA (siRNA).

The term "short interfering RNA" or "siRNA" as used herein refers to a double stranded nucleic acid molecule capable of RNA interference "RNAi", see for example Bass, 2001, *Nature*, 411, 428-429; Elbashir et al., 2001, *Nature*, 411, 494-498; and Kreutzer *et al.*, International PCT Publication No. WO 00/44895; Zernicka-Goetz *et al.*, International PCT Publication No. WO 01/36646; Fire, International PCT Publication No. WO 99/32619; Plaetinck *et al.*, International PCT Publication No. WO 00/01846; Mello and Fire, International PCT Publication No. WO 01/29058; Deschamps-Depaillette, International PCT Publication No. WO 99/07409; and Li *et al.*, International PCT Publication No. WO 00/44914. As used herein, siRNA molecules need not be limited to those molecules containing only RNA, but further encompasses chemically modified nucleotides and non-nucleotides.

The term "allozyme" as used herein refers to an allosteric enzymatic nucleic acid molecule, see for example see for example George et al., US Patent Nos. 5,834,186 and 5,741,679, Shih et al., US Patent No. 5,589,332, Nathan et al., US Patent No 5,871,914, Nathan and Ellington, International PCT publication No. WO 00/24931, Breaker et al., International PCT Publication Nos. WO 00/26226 and 98/27104, and Sullenger et al., International PCT publication No. WO 99/29842.

The term "cell" as used herein, refers to its usual biological sense, and does not refer to an entire multicellular organism. The cell can, for example, be in vitro, e.g., in cell culture, or present in a multicellular organism, including,, e.g., birds, plants and mammals such as humans, cows, sheep, apes, monkeys, swine, dogs, and cats. The cell can be prokaryotic (e.g., bacterial cell) or eukaryotic (e.g., mammalian or plant cell).

The term "highly conserved sequence region" as used herein, refers to a nucleotide sequence of one or more regions in a target gene does not vary significantly from one generation to the other or from one biological system to the other.

The term "non-nucleotide" as used herein, refers to any group or compound which can be incorporated into a nucleic acid chain in the place of one or more nucleotide units, including either sugar and/or phosphate substitutions, and allows the remaining bases to exhibit their enzymatic activity. The group or compound is abasic in that it does not contain a commonly recognized nucleotide base, such as adenosine, guanine, cytosine, uracil or thymine.

The term "nucleotide" as used herein, refers to a heterocyclic nitrogenous base in N-glycosidic linkage with a phosphorylated sugar. Nucleotides are recognized in the art to include natural bases (standard), and modified bases well known in the art. Such bases are generally located at the 1' position of a nucleotide sugar moiety. Nucleotides generally comprise a base, sugar and a phosphate group. The nucleotides can be unmodified or modified at the sugar, phosphate and/or base moiety, (also referred to interchangeably as nucleotide analogs, modified nucleotides, non-natural nucleotides, non-standard nucleotides and other; see for example, Usman and McSwiggen, supra; Eckstein *et al.*, International PCT Publication No. WO 92/07065; Usman *et al.*, International PCT Publication No. WO 93/15187; Uhlman & Peyman, supra all are hereby incorporated by reference herein). There are several examples of modified nucleic acid bases known in the art as summarized by Limbach *et al.*, 1994, Nucleic Acids Res. 22, 2183. Some of the non-limiting examples of chemically modified and other natural nucleic acid bases that can be introduced into nucleic acids include, for example, inosine, purine, pyridin-4-one, pyridin-2-one, phenyl, pseudouracil, 2, 4, 6-trimethoxy benzene, 3-methyl uracil, dihydrouridine, naphthyl, aminophenyl, 5-alkylcytidines (e.g., 5-methylcytidine), 5-alkyluridines (e.g., ribothymidine), 5-halouridine (e.g., 5-bromouridine) or 6-azapyrimidines or 6-alkylpyrimidines (e.g. 6-methyluridine), propyne, quesosine, 2-thiouridine, 4-thiouridine, wybutosine, wybutoxosine, 4-acetylcytidine, 5-(carboxyhydroxymethyl)uridine, 5'-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluridine, beta-D-galactosylqueosine, 1-methyladenosine, 1-methylinosine, 2,2-dimethylguanosine, 3-methylcytidine, 2-methyladenosine, 2-methylguanosine, N6-methyladenosine, 7-methylguanosine, 5-methoxyaminomethyl-2-thiouridine, 5-methylaminomethyluridine, 5-methylcarbonylmethyluridine, 5-methyloxyuridine, 5-methyl-2-thiouridine, 2-methylthio-N6-isopentenyladenosine, beta-D-mannosylqueosine, uridine-5-oxyacetic acid, 2-thiocytidine, threonine derivatives and others (Burgin *et al.*, 1996, Biochemistry, 35, 14090; Uhlman & Peyman, supra). By "modified bases" in this aspect is meant nucleotide bases other than adenine, guanine, cytosine and uracil at 1' position or their equivalents; such bases can be used at any position, for example, within the catalytic core of an enzymatic nucleic acid molecule and/or in the substrate-binding regions of the nucleic acid molecule.

The term "nucleoside" as used herein, refers to a heterocyclic nitrogenous base in N-glycosidic linkage with a sugar. Nucleosides are recognized in the art to include natural bases (standard), and modified bases well known in the art. Such bases are generally located at the 1' position of a nucleoside sugar moiety. Nucleosides generally comprise a base and sugar group. The nucleosides can be unmodified or modified at the sugar, and/or base moiety, (also referred to interchangeably as nucleoside analogs, modified nucleosides, non-natural nucleosides, non-standard nucleosides and other; see for example, Usman and McSwiggen, supra; Eckstein *et al.*, International PCT Publication No. WO 92/07065; Usman *et al.*, International PCT Publication No. WO 93/15187; Uhlman & Peyman, supra all are hereby incorporated by reference herein). There are several examples of modified nucleic acid bases known in the art as summarized by Limbach *et al.*, 1994, Nucleic Acids Res. 22, 2183. Some of the non-limiting examples of chemically modified and other natural nucleic acid bases that can be introduced into nucleic acids include, inosine, purine, pyridin-4-one, pyridin-2-one, phenyl, pseudouracil, 2, 4, 6-trimethoxy benzene, 3-methyl uracil, dihydrouridine, naphthyl, aminophenyl, 5-alkylcytidines (e.g., 5-methylcytidine), 5-alkyluridines (e.g., ribothymidine), 5-halouridine (e.g., 5-bromouridine) or 6-azapyrimidines or 6-alkylpyrimidines (e.g. 6-methyluridine), propyne, quesosine, 2-thiouridine, 4-thiouridine, wybutosine, wybutoxosine, 4-acetylcytidine, 5-(carboxyhydroxymethyl)uridine, 5'-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluridine, beta-D-galactosylqueosine, 1-methyladenosine, 1-methylinosine, 2,2-dimethylguanosine, 3-methylcytidine, 2-methyladenosine, 2-methylguanosine, N6-methyladenosine, 7-methylguanosine, 5-methoxyaminomethyl-2-thiouridine, 5-methylaminomethyluridine, 5-methylcarbonylmethyluridine, 5-methyloxyuridine, 5-methyl-2-thiouridine, 2-methylthio-N6-isopentenyladenosine, beta-D-mannosylqueosine, uridine-5-oxyacetic acid, 2-thiocytidine, threonine derivatives and others (Burgin *et al.*, 1996, Biochemistry, 35, 14090; Uhlman & Peyman, supra). By "modified bases" in this aspect is meant nucleoside bases other than adenine, guanine, cytosine and uracil at 1' position or their equivalents; such bases can be used at any position, for example, within the catalytic core of an enzymatic nucleic acid molecule and/or in the substrate-binding regions of the nucleic acid molecule.

The term "cap structure" as used herein, refers to chemical modifications, which have been incorporated at either terminus of the oligonucleotide (see for example Wincott *et al.*, WO 97/26270, incorporated by reference herein). These terminal modifications protect the nucleic acid molecule from exonuclease degradation, and can help in delivery and/or localization within a cell. The cap can be present at the 5'-terminus (5'-cap) or at the 3'-terminus (3'-cap) or can be present on both terminus. In non-limiting examples,

the 5'-cap includes inverted abasic residue (moiety), 4',5'-methylene nucleotide; 1-(beta-D-erythrofuranosyl) nucleotide, 4'-thio nucleotide, carbocyclic nucleotide; 1,5-anhydrohexitol nucleotide; L-nucleotides; alpha-nucleotides; modified base nucleotide; phosphorodithioate linkage; *threo*-pentofuranosyl nucleotide; acyclic 3',4'-seco

5   nucleotide; acyclic 3,4-dihydroxybutyl nucleotide; acyclic 3,5-dihydroxypentyl nucleotide, 3'-3'-inverted nucleotide moiety; 3'-3'-inverted abasic moiety; 3'-2'-inverted nucleotide moiety; 3'-2'-inverted abasic moiety; 1,4-butanediol phosphate; 3'-phosphoramidate; hexylphosphate; aminohexyl phosphate; 3'-phosphate; 3'-phosphorothioate; phosphorodithioate; or bridging or non-bridging methylphosphonate

10   moiety (for more details see Wincott *et al.,* International PCT publication No. WO 97/26270, incorporated by reference herein).

The term "abasic" as used herein, refers to sugar moieties lacking a base or having other chemical groups in place of a base at the 1' position, for example a 3',3'-linked or 5',5'-linked deoxyabasic ribose derivative (for more details see Wincott *et al.,*

15   International PCT publication No. WO 97/26270).

The term "unmodified nucleoside" as used herein, refers to one of the bases adenine, cytosine, guanine, thymine, uracil joined to the 1' carbon of β-D-ribo-furanose.

The term "modified nucleoside" as used herein, refers to any nucleotide base which contains a modification in the chemical structure of an unmodified nucleotide base, sugar

20   and/or phosphate.

The term "consists essentially of" as used herein, is meant that the active nucleic acid molecule of the invention, for example, an enzymatic nucleic acid molecule, contains an enzymatic center or core equivalent to those in the examples, and binding arms able to bind RNA such that cleavage at the target site occurs. Other sequences can be present

25   which do not interfere with such cleavage. Thus, a core region can, for example, include one or more loop, stem-loop structure, or linker which does not prevent enzymatic activity. For example, a core sequence for a hammerhead enzymatic nucleic acid can comprise a conserved sequence, such as 5'-CUGAUGAG-3' and 5'-CGAA-3' connected by "X", where X is 5'-GCCGUUAGGC-3' (SEQ ID NO 1), or any other Stem II region

30   known in the art, or a nucleotide and/or non-nucleotide linker. Similarly, for other nucleic acid molecules of the instant invention, such as Inozyme, G-cleaver, amberzyme, zinzyme, DNAzyme, antisense, 2-5A antisense, triplex forming nucleic acid, and decoy nucleic acids, other sequences or non-nucleotide linkers can be present that do not interfere with the function of the nucleic acid molecule.

Sequence X can be a linker of ≥ 2 nucleotides in length, preferably 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 26, 30, where the nucleotides can preferably be internally base-paired to form a stem of preferably ≥ 2 base pairs. In yet another embodiment, the nucleotide linker X can be a nucleic acid aptamer, such as an ATP aptamer, HIV Rev aptamer (RRE), HIV

5   Tat aptamer (TAR) and others (for a review see Gold et al., 1995, Annu. Rev. Biochem., 64, 763; and Szostak & Ellington, 1993, in The RNA World, ed. Gesteland and Atkins, pp. 511, CSH Laboratory Press). A "nucleic acid aptamer" as used herein is meant to indicate a nucleic acid sequence capable of interacting with a ligand. The ligand can be any natural or a synthetic molecule, including but not limited to a resin, metabolites,

10   nucleosides, nucleotides, drugs, toxins, transition state analogs, peptides, lipids, proteins, amino acids, nucleic acid molecules, hormones, carbohydrates, receptors, cells, viruses, bacteria and others.

Alternatively or in addition, sequence X can be a non-nucleotide linker. Non-nucleotides can include abasic nucleotide, polyether, polyamine, polyamide, peptide,

15   carbohydrate, lipid, or polyhydrocarbon compounds. Specific examples include those described by Seela and Kaiser, Nucleic Acids Res. 1990, 18:6353 and Nucleic Acids Res. 1987, 15:3113; Cload and Schepartz, J. Am. Chem. Soc. 1991, 113:6324; Richardson and Schepartz, J. Am. Chem. Soc. 1991, 113:5109; Ma et al., Nucleic Acids Res. 1993, 21:2585 and Biochemistry 1993, 32:1751; Durand et al., Nucleic Acids Res. 1990,

20   18:6353; McCurdy et al., Nucleosides & Nucleotides 1991, 10:287; Jschke et al., Tetrahedron Lett. 1993, 34:301; Ono et al., Biochemistry 1991, 30:9914; Arnold et al., International Publication No. WO 89/02439; Usman et al., International Publication No. WO 95/06731; Dudycz et al., International Publication No. WO 95/11910 and Ferentz and Verdine, J. Am. Chem. Soc. 1991, 113:4000, all hereby incorporated by reference

25   herein. A "non-nucleotide" further means any group or compound which can be incorporated into a nucleic acid chain in the place of one or more nucleotide units, including either sugar and/or phosphate substitutions, and allows the remaining bases to exhibit their enzymatic activity. The group or compound can be abasic in that it does not contain a commonly recognized nucleotide base, such as adenosine, guanine, cytosine,

30   uracil or thymine. Thus, in a preferred embodiment, the invention features an enzymatic nucleic acid molecule having one or more non-nucleotide moieties, and having enzymatic activity to cleave an RNA or DNA molecule.

The term "patient" as used herein, refers to an organism, which is a donor or recipient of explanted cells or the cells themselves. "Patient" also refers to an organism to

35   which the nucleic acid molecules of the invention can be administered. Preferably, a

patient is a mammal or mammalian cells. More preferably, a patient is a human or human cells.

The term "enhanced enzymatic activity" as used herein, includes activity measured in cells and/or in vivo where the activity is a reflection of both the catalytic activity and the stability of the nucleic acid molecules of the invention. In this invention, the product of these properties can be increased in vivo compared to an all RNA enzymatic nucleic acid or all DNA enzyme. In some cases, the activity or stability of the nucleic acid molecule can be decreased (i.e., less than ten-fold), but the overall activity of the nucleic acid molecule is enhanced, in vivo.

By "comprising" is meant including, but not limited to, whatever follows the word "comprising". Thus, use of the term "comprising" indicates that the listed elements are required or mandatory, but that other elements are optional and can or can not be present. By "consisting of" is meant including, and limited to, whatever follows the phrase "consisting of". Thus, the phrase "consisting of" indicates that the listed elements are required or mandatory, and that no other elements can be present.

The term "negatively charged molecules" as used herein, refers to molecules such as nucleic acid molecules (e.g., RNA, DNA, oligonucleotides, mixed polymers, peptide nucleic acid, and the like), peptides (e.g., polyaminoacids, polypeptides, proteins and the like), nucleotides, pharmaceutical and biological compositions, that have negatively charged groups that can ion-pair with the positively charged head group of the cationic lipids of the invention.

The term "coupling" as used herein, refers to a reaction, either chemical or enzymatic, in which one atom, moiety, group, compound or molecule is joined to another atom, moiety, group, compound or molecule.

The terms "deprotection" or "deprotecting" as used herein, refers to the removal of a protecting group.

The term "alkyl" as used herein refers to a saturated aliphatic hydrocarbon, including straight-chain, branched-chain "isoalkyl", and cyclic alkyl groups. The term "alkyl" also comprises alkoxy, alkyl-thio, alkyl-thio-alkyl, alkoxyalkyl, alkylamino, alkenyl, alkynyl, alkoxy, cycloalkenyl, cycloalkyl, cycloalkylalkyl, heterocycloalkyl, heteroaryl, C1-C6 hydrocarbyl, aryl or substituted aryl groups. Preferably, the alkyl group has 1 to 12 carbons. More preferably it is a lower alkyl of from about 1 to about 7 carbons, more preferably about 1 to about 4 carbons. The alkyl group can be substituted

or unsubstituted. When substituted the substituted group(s) preferably comprise hydroxy, oxy, thio, amino, nitro, cyano, alkoxy, alkyl-thio, alkyl-thio-alkyl, alkoxyalkyl, alkylamino, silyl, alkenyl, alkynyl, alkoxy, cycloalkenyl, cycloalkyl, cycloalkylalkyl, heterocycloalkyl, heteroaryl, C1-C6 hydrocarbyl, aryl or substituted aryl groups. The term "alkyl" also includes alkenyl groups containing at least one carbon-carbon double bond, including straight-chain, branched-chain, and cyclic groups. Preferably, the alkenyl group has about 2 to about 12 carbons. More preferably it is a lower alkenyl of from about 2 to about 7 carbons, more preferably about 2 to about 4 carbons. The alkenyl group can be substituted or unsubstituted. When substituted the substituted group(s) preferably comprise hydroxy, oxy, thio, amino, nitro, cyano, alkoxy, alkyl-thio, alkyl-thio-alkyl, alkoxyalkyl, alkylamino, silyl, alkenyl, alkynyl, alkoxy, cycloalkenyl, cycloalkyl, cycloalkylalkyl, heterocycloalkyl, heteroaryl, C1-C6 hydrocarbyl, aryl or substituted aryl groups. The term "alkyl" also includes alkynyl groups containing at least one carbon-carbon triple bond, including straight-chain, branched-chain, and cyclic groups. Preferably, the alkynyl group has about 2 to about 12 carbons. More preferably it is a lower alkynyl of from about 2 to about 7 carbons, more preferably about 2 to about 4 carbons. The alkynyl group can be substituted or unsubstituted. When substituted the substituted group(s) preferably comprise hydroxy, oxy, thio, amino, nitro, cyano, alkoxy, alkyl-thio, alkyl-thio-alkyl, alkoxyalkyl, alkylamino, silyl, alkenyl, alkynyl, alkoxy, cycloalkenyl, cycloalkyl, cycloalkylalkyl, heterocycloalkyl, heteroaryl, C1-C6 hydrocarbyl, aryl or substituted aryl groups. Alkyl groups or moieties of the invention can also include aryl, alkylaryl, carbocyclic aryl, heterocyclic aryl, amide and ester groups. The preferred substituent(s) of aryl groups are halogen, trihalomethyl, hydroxyl, SH, OH, cyano, alkoxy, alkyl, alkenyl, alkynyl, and amino groups. An "alkylaryl" group refers to an alkyl group (as described above) covalently joined to an aryl group (as described above). Carbocyclic aryl groups are groups wherein the ring atoms on the aromatic ring are all carbon atoms. The carbon atoms are optionally substituted. Heterocyclic aryl groups are groups having from about 1 to about 3 heteroatoms as ring atoms in the aromatic ring and the remainder of the ring atoms are carbon atoms. Suitable heteroatoms include oxygen, sulfur, and nitrogen, and include furanyl, thienyl, pyridyl, pyrrolyl, N-lower alkyl pyrrolo, pyrimidyl, pyrazinyl, imidazolyl and the like, all optionally substituted. An "amide" refers to an -C(O)-NH-R, where R is either alkyl, aryl, alkylaryl or hydrogen. An "ester" refers to an -C(O)-OR', where R is either alkyl, aryl, alkylaryl or hydrogen.

The term "alkoxyalkyl" as used herein refers to an alkyl-O-alkyl ether, for example, methoxyethyl or ethoxymethyl.

The term "alkyl-thio-alkyl" as used herein refers to an alkyl-S-alkyl thioether, for example, methylthiomethyl or methylthioethyl.

The term "amino" as used herein refers to a nitrogen containing group as is known in the art derived from ammonia by the replacement of one or more hydrogen radicals by organic radicals. For example, the terms "aminoacyl" and "aminoalkyl" refer to specific N-substituted organic radicals with acyl and alkyl substituent groups respectively.

The term "amination" as used herein refers to a process in which an amino group or substituted amine is introduced into an organic molecule.

The term "exocyclic amine protecting moiety" as used herein refers to a nucleobase amino protecting group compatible with oligonucleotide synthesis, for example, an acyl or amide group.

The term "alkenyl" as used herein refers to a straight or branched hydrocarbon of a designed number of carbon atoms containing at least one carbon-carbon double bond. Examples of "alkenyl" include vinyl, allyl, and 2-methyl-3-heptene.

The term "alkoxy" as used herein refers to an alkyl group of indicated number of carbon atoms attached to the parent molecular moiety through an oxygen bridge. Examples of alkoxy groups include, for example, methoxy, ethoxy, propoxy and isopropoxy.

The term "alkynyl" as used herein refers to a straight or branched hydrocarbon of a designed number of carbon atoms containing at least one carbon-carbon triple bond. Examples of "alkynyl" include propargyl, propyne, and 3-hexyne.

The term "aryl" as used herein refers to an aromatic hydrocarbon ring system containing at least one aromatic ring. The aromatic ring can optionally be fused or otherwise attached to other aromatic hydrocarbon rings or non-aromatic hydrocarbon rings. Examples of aryl groups include, for example, phenyl, naphthyl, 1,2,3,4-tetrahydronaphthalene and biphenyl. Preferred examples of aryl groups include phenyl and naphthyl.

The term "cycloalkenyl" as used herein refers to a C3-C8 cyclic hydrocarbon containing at least one carbon-carbon double bond. Examples of cycloalkenyl include cyclopropenyl, cyclobutenyl, cyclopentenyl, cyclopentadiene, cyclohexenyl, 1,3-cyclohexadiene, cycloheptenyl, cycloheptatrienyl, and cyclooctenyl.

The term "cycloalkyl" as used herein refers to a C3-C8 cyclic hydrocarbon. Examples of cycloalkyl include cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cycloheptyl and cyclooctyl.

5   The term "cycloalkylalkyl," as used herein, refers to a C3-C7 cycloalkyl group attached to the parent molecular moiety through an alkyl group, as defined above. Examples of cycloalkylalkyl groups include cyclopropylmethyl and cyclopentylethyl.

The terms "halogen" or "halo" as used herein refers to indicate fluorine, chlorine, bromine, and iodine.

The term "heterocycloalkyl," as used herein refers to a non-aromatic ring system
10   containing at least one heteroatom selected from nitrogen, oxygen, and sulfur. The heterocycloalkyl ring can be optionally fused to or otherwise attached to other heterocycloalkyl rings and/or non-aromatic hydrocarbon rings. Preferred heterocycloalkyl groups have from 3 to 7 members. Examples of heterocycloalkyl groups include, for example, piperazine, morpholine, piperidine, tetrahydrofuran, pyrrolidine, and pyrazole.
15   Preferred heterocycloalkyl groups include piperidinyl, piperazinyl, morpholinyl, and pyrolidinyl.

The term "heteroaryl" as used herein refers to an aromatic ring system containing at least one heteroatom selected from nitrogen, oxygen, and sulfur. The heteroaryl ring can be fused or otherwise attached to one or more heteroaryl rings, aromatic or non-
20   aromatic hydrocarbon rings or heterocycloalkyl rings. Examples of heteroaryl groups include, for example, pyridine, furan, thiophene, 5,6,7,8-tetrahydroisoquinoline and pyrimidine. Preferred examples of heteroaryl groups include thienyl, benzothienyl, pyridyl, quinolyl, pyrazinyl, pyrimidyl, imidazolyl, benzimidazolyl, furanyl, benzofuranyl, thiazolyl, benzothiazolyl, isoxazolyl, oxadiazolyl, isothiazolyl, benzisothiazolyl, triazolyl,
25   tetrazolyl, pyrrolyl, indolyl, pyrazolyl, and benzopyrazolyl.

The term "C1-C6 hydrocarbyl" as used herein refers to straight, branched, or cyclic alkyl groups having 1-6 carbon atoms, optionally containing one or more carbon-carbon double or triple bonds. Examples of hydrocarbyl groups include, for example, methyl, ethyl, propyl, isopropyl, n-butyl, sec-butyl, tert-butyl, pentyl, 2-pentyl, isopentyl,
30   neopentyl, hexyl, 2-hexyl, 3-hexyl, 3-methylpentyl, vinyl, 2-pentene, cyclopropylmethyl, cyclopropyl, cyclohexylmethyl, cyclohexyl and propargyl. When reference is made herein to C1-C6 hydrocarbyl containing one or two double or triple bonds it is understood that at least two carbons are present in the alkyl for one double or triple bond, and at least four carbons for two double or triple bonds.

The term "protecting group" as used herein, refers to groups known in the art that are readily introduced and removed from an atom, for example O, N, P, or S. Protecting groups are used to prevent undesirable reactions from taking place that can compete with the formation of a specific compound or intermediate of interest. See also "Protective Groups in Organic Synthesis", 3rd Ed., 1999, Greene, T. W. and related publications.

The term "nitrogen protecting group," as used herein, refers to groups known in the art that are readily introduced on to and removed from a nitrogen. Examples of nitrogen protecting groups include Boc, Cbz, benzoyl, and benzyl. See also "Protective Groups in Organic Synthesis", 3rd Ed., 1999, Greene, T. W. and related publications.

The term "hydroxy protecting group," or "hydroxy protection" as used herein, refers to groups known in the art that are readily introduced on to and removed from an oxygen, specifically an –OH group. Examples of hyroxy protecting groups include trityl or substituted trityl goups, such as monomethoxytrityl and dimethoxytrityl, or substituted silyl groups, such as tert-butyldimethyl, trimethylsilyl, or tert-butyldiphenyl silyl groups. See also "Protective Groups in Organic Synthesis", 3rd Ed., 1999, Greene, T. W. and related publications.

The term "acyl" as used herein refers to -C(O)R groups, wherein R is an alkyl or aryl.

The term "phosphorus containing group" as used herein, refers to a chemical group containing a phosphorus atom. The phosphorus atom can be trivalent or pentavalent, and can be substituted with O, H, N, S, C or halogen atoms. Examples of phosphorus containing groups of the instant invention include but are not limited to phosphorus atoms substituted with O, H, N, S, C or halogen atoms, comprising phosphonate, alkylphosphonate, phosphate, diphosphate, triphosphate, pyrophosphate, phosphorothioate, phosphorodithioate, phosphoramidate, phosphoramidite groups, nucleotides and nucleic acid molecules.

The term "phosphine" or "phosphite" as used herein refers to a trivalent phosphorus species, for example compounds having Formula 97:

$$\begin{array}{c} R \diagdown \ \diagup S \\ P \\ | \\ T \end{array}$$
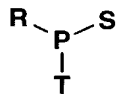
wherein R can include the groups:

ξ—CH$_3$ CH$_3$O—ξ N≡C$\diagup$$\diagdown$$\diagup$O$\diagdown$$\xi$ N≡C$\diagup$$\diagdown$$\diagup$S$\diagdown$$\xi$

[benzene ring]—S—$\diagup$$\diagdown$—O—$\xi$ or   Cl—[benzene ring with Cl]—S—$\diagup$$\diagdown$—O—$\xi$
                    ‖   ‖
                    O   O
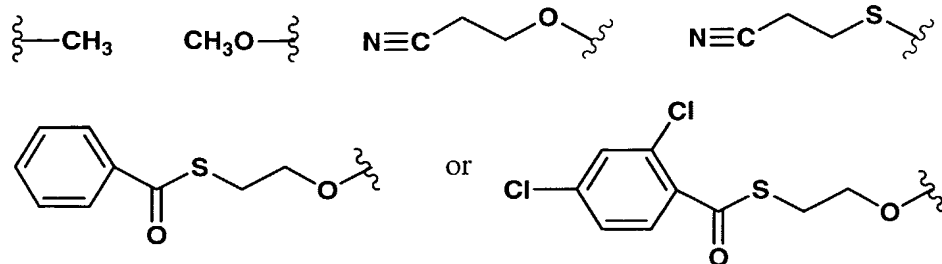
and wherein S and T independently include the groups:

ξ—N$\diagup$$\diagdown$ ξ—N$\diagup^{CH_2CH_3}_{CH_2CH_3}$ ξ—N$\diagup$$\diagdown$ ξ—N$\bigcirc$ or ξ—N$\bigcirc$O
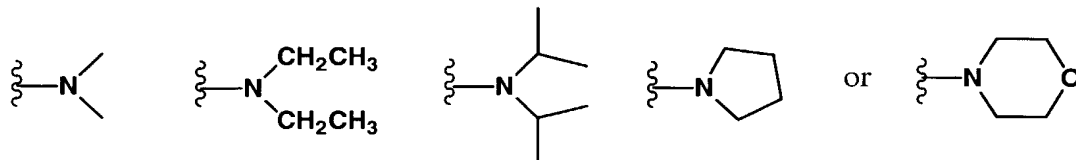
The term "phosphate" as used herein refers to a pentavalent phosphorus species, for example a compound having Formula 98:

$$\begin{array}{c} S \\ \| \\ R—P=M \\ \| \\ T \end{array}$$
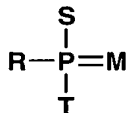
wherein R includes the groups:

ξ—CH$_3$ CH$_3$O—ξ N≡C$\diagup$$\diagdown$$\diagup$O$\diagdown$$\xi$ N≡C$\diagup$$\diagdown$$\diagup$S$\diagdown$$\xi$

[benzene ring]—S—$\diagup$$\diagdown$—O—$\xi$ or   Cl—[benzene ring with Cl]—S—$\diagup$$\diagdown$—O—$\xi$
                    ‖   ‖
                    O   O
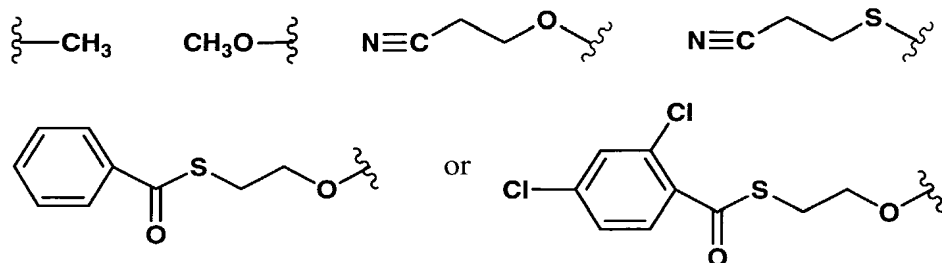
and wherein S and T each independently can be a sulfur or oxygen atom or a group which can include:

ξ—N$\diagup$$\diagdown$ ξ—N$\diagup^{CH_2CH_3}_{CH_2CH_3}$ ξ—N$\diagup$$\diagdown$ ξ—N$\bigcirc$ or ξ—N$\bigcirc$O
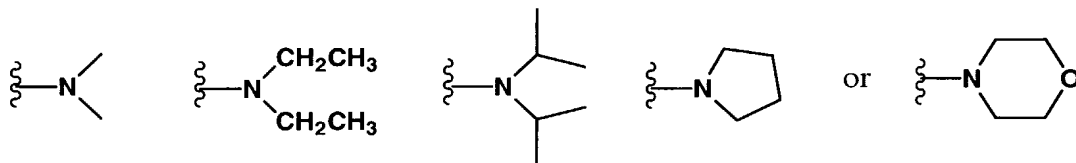
and wherein M comprises a sulfur or oxygen atom. The phosphate of the invention can comprise a nucleotide phosphate, wherein any R, S, or T in Formula 98 comprises a linkage to a nucleic acid or nucleoside.

The term "cationic salt" as used herein refers to any organic or inorganic salt having a net positive charge, for example a triethylammonium (TEA) salt.

The term "degradable linker" as used herein, refers to linker moieties that are capable of cleavage under various conditions. Conditions suitable for cleavage can include but are not limited to pH, UV irradiation, enzymatic activity, temperature, hydrolysis, elimination, and substitution reactions, and thermodynamic properties of the linkage.

The term "photolabile linker" as used herein, refers to linker moieties as are known in the art, that are selectively cleaved under particular UV wavelengths. Compounds of the invention containing photolabile linkers can be used to deliver compounds to a target cell or tissue of interest, and can be subsequently released in the presence of a UV source.

The term "nucleic acid conjugates" as used herein, refers to nucleoside, nucleotide and oligonucleotide conjugates.

The term "folate" as used herein, refers to analogs and derivatives of folic acid, for example antifolates, dihydrofloates, tetrahydrofolates, tetrahydorpterins, folinic acid, pteropolyglutamic acid, 1-deza, 3-deaza, 5-deaza, 8-deaza, 10-deaza, 1,5-deaza, 5,10 dideaza, 8,10-dideaza, and 5,8-dideaza folates, antifolates, and pteroic acid derivatives.

The term "compounds with neutral charge" as used herein, refers to compositions which are neutral or uncharged at neutral or physiological pH. Examples of such compounds are cholesterol and other steroids, cholesteryl hemisuccinate (CHEMS), dioleoyl phosphatidyl choline, distearoylphosphotidyl choline (DSPC), fatty acids such as oleic acid, phosphatidic acid and its derivatives, phosphatidyl serine, polyethylene glycol - conjugated phosphatidylamine, phosphatidylcholine, phosphatidylethanolamine and related variants, prenylated compounds including farnesol, polyprenols, tocopherol, and their modified forms, diacylsuccinyl glycerols, fusogenic or pore forming peptides, dioleoylphosphotidylethanolamine (DOPE), ceramide and the like.

The term "lipid aggregate" as used herein refers to a lipid-containing composition wherein the lipid is in the form of a liposome, micelle (non-lamellar phase) or other aggregates with one or more lipids.

The term "biological system" as used herein, refers to a eukaryotic system or a prokaryotic system, can be a bacterial cell, plant cell or a mammalian cell, or can be of plant origin, mammalian origin, yeast origin, Drosophila origin, or archebacterial origin.

The term "systemic administration" as used herein refers to the in vivo systemic absorption or accumulation of drugs in the blood stream followed by distribution throughout the entire body. Administration routes which lead to systemic absorption

include, without limitations: intravenous, subcutaneous, intraperitoneal, inhalation, oral, intrapulmonary and intramuscular. Each of these administration routes expose the desired negatively charged polymers, e.g., nucleic acids, to an accessible diseased tissue. The rate of entry of a drug into the circulation has been shown to be a function of molecular

5   weight or size. The use of a liposome or other drug carrier comprising the compounds of the instant invention can potentially localize the drug, for example, in certain tissue types, such as the tissues of the reticular endothelial system (RES). A liposome formulation which can facilitate the association of drug with the surface of cells, such as, lymphocytes and macrophages is also useful. This approach can provide enhanced delivery of the drug

10   to target cells by taking advantage of the specificity of macrophage and lymphocyte immune recognition of abnormal cells, such as the cancer cells.

The term "pharmacological composition" or "pharmaceutical formulation" refers to a composition or formulation in a form suitable for administration, for example, systemic administration, into a cell or patient, preferably a human. Suitable forms, in

15   part, depend upon the use or the route of entry, for example oral, transdermal, or by injection. Such forms should not prevent the composition or formulation to reach a target cell (i.e., a cell to which the negatively charged polymer is targeted).

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

20   <u>Description of the Preferred Embodiments</u>

The drawings will be first described briefly.

<u>Drawings:</u>

Figure 1 shows examples of chemically stabilized ribozyme motifs. HH Rz, represents hammerhead ribozyme motif (Usman et al., 1996, Curr. Op. Struct. Bio., 1,

25   527); NCH Rz represents the NCH ribozyme motif (Ludwig & Sproat, International PCT Publication No. WO 98/58058); G-Cleaver, represents G-cleaver ribozyme motif (Kore et al., 1998, Nucleic Acids Research 26, 4116-4120, Eckstein et al., International PCT publication No. WO 99/16871). N or n, represent independently a nucleotide which can be same or different and have complementarity to each other; rI, represents ribo-Inosine

30   nucleotide; arrow indicates the site of cleavage within the target. Position 4 of the HH Rz and the NCH Rz is shown as having 2'-C-allyl modification, but those skilled in the art will recognize that this position can be modified with other modifications well known in

the art, so long as such modifications do not significantly inhibit the activity of the ribozyme.

Figure 2 shows an example of the Amberzyme ribozyme motif that is chemically stabilized (see for example Beigelman et al., International PCT publication No. WO 99/55857).

Figure 3 shows an example of the Zinzyme A ribozyme motif that is chemically stabilized (see for example Beigelman et al., Beigelman et al., International PCT publication No. WO 99/55857).

Figure 4 shows an example of a DNAzyme motif described by Santoro et al., 1997, PNAS, 94, 4262.

Figure 5 shows a synthetic scheme for the synthesis of a folate conjugate of the instant invention.

Figure 6 shows representative examples of fludarabine-folate conjugate molecules of the invention.

Figure 7 shows a synthetic scheme for post-synthetic modification of a nucleic acid molecule to produce a folate conjugate.

Figure 8 shows a synthetic scheme for generating a protected pteroic acid synthon of the invention.

Figure 9 shows a synthetic scheme for generating a 2-dithiopyridyl activated folic acid synthon of the invention.

Figure 10 shows a synthetic scheme for generating an oligonucleotide or nucleic acid-folate conjugate.

Figure 11 shows a an alternative synthetic scheme for generating an oligonucleotide or nucleic acid-folate conjugate.

Figure 12 shows an alternative synthetic scheme for post-synthetic modification of a nucleic acid molecule to produce a folate conjugate.

Figure 13 shows a non-limiting example of a synthetic scheme for the synthesis of a N-acetyl-D-galactosamine-2'-aminouridine phosphoramidite conjugate of the invention.

Figure 14 shows a non-limiting example of a synthetic scheme for the synthesis of a N-acetyl-D-galactosamine-D-threoninol phosphoramidite conjugate of the invention.

Figure 15 shows a non-limiting example of an N-acetyl-D-galactosamine enzymatic nucleic acid conjugate of the invention. **W** shown in the example refers to a biodegradable linker, for example a nucleic acid dimer, trimer, or tetramer comprising ribonucleotides and/or deoxyribonucleotides.

Figure 16 shows a non-limiting example of a synthetic scheme for the synthesis of a dodecanoic acid derived conjugate linker of the invention.

Figure 17 shows a non-limiting example of a synthetic scheme for the synthesis of an oxime linked nucleic acid/peptide conjugate of the invention.

Figure 18 shows non-limiting examples of phospholipid derived nucleic acid conjugates of the invention. **W** shown in the examples refers to a biodegradable linker, for example a nucleic acid dimer, trimer, or tetramer comprising ribonucleotides and/or deoxyribonucleotides.

Figure 19 shows a non-limiting example of a synthetic scheme for preparing a phospholipid derived enzymatic nucleic acid conjugates of the invention.

Figure 20 shows a non-limiting example of a synthetic scheme for preparing a polyethylene glycol (PEG) derived enzymatic nucleic acid conjugates of the invention.

Figure 21 shows PK data of a 40K PEG conjugated enzymatic nucleic acid molecule compared to the corresponding non-conjugated enzymatic nucleic acid molecule. The graph is a time course of serum concentration in mice dosed with 30 mg/kg of Angiozyme™ or 40-kDa-PEG-Angiozyme™. The hybridization method was used to quantitate Angiozyme™ levels.

Figure 22 shows PK data of a phospholipid conjugated enzymatic nucleic acid molecule compared to the corresponding non-conjugated enzymatic nucleic acid molecule.

Figure 23 shows a non-limiting example of a synthetic scheme for preparing a poly-N-acetyl-D-galactosamine enzymatic nucleic acid conjugate of the invention.

Figure 24a-b shows a non-limiting example of a synthetic approach for synthesizing peptide or protein conjugates to PEG utilizing a biodegradable linker using oxime and morpholino linkages.

Figure 25 shows a non-limiting example of a synthetic approach for synthesizing peptide or protein conjugates to PEG utilizing a biodegradable linker using oxime and phosphoramidate linkages.

Figure 26a-b shows a non-limiting example of a synthetic approach for synthesizing peptide or protein conjugates to PEG utilizing a biodegradable linker using phosphoramidate linkages.

Figure 27 shows non-limiting examples of phospholipid derived protein/peptide conjugates of the invention. **W** shown in the examples refers to a biodegradable linker, for example a nucleic acid dimer, trimer, or tetramer comprising ribonucleotides and/or deoxyribonucleotides.

Figure 28 shows a non-limiting example of an N-acetyl-D-galactosamine peptide/protein conjugate of the invention, the example shown is with a peptide. **W** shown in the example refers to a biodegradable linker, for example a nucleic acid dimer, trimer, or tetramer comprising ribonucleotides and/or deoxyribonucleotides.

Figure 29 shows a non-limiting example of a synthetic approach for synthesizing peptide or protein conjugates to PEG utilizing a biodegradable linker using phosphoramidate linkages via coupling a protein phosphoramidite to a PEG conjugated nucleic acid linker.

Method of Use

The compositions and conjugates of the instant invention can be used to administer pharmaceutical agents. Pharmaceutical agents prevent, inhibit the occurrence, or treat (alleviate a symptom to some extent, preferably all of the symptoms) of a disease state in a patient.

Generally, the compounds of the instant invention are introduced by any standard means, with or without stabilizers, buffers, and the like, to form a pharmaceutical composition. For use of a liposome delivery mechanism, standard protocols for formation of liposomes can be followed. The compositions of the present invention can also be formulated and used as tablets, capsules or elixirs for oral administration; suppositories for rectal administration; sterile solutions; suspensions for injectable administration; and the like.

The present invention also includes pharmaceutically acceptable formulations of the compounds described above, preferably in combination with the molecule(s) to be

delivered. These formulations include salts of the above compounds, e.g., acid addition salts, for example, salts of hydrochloric, hydrobromic, acetic acid, and benzene sulfonic acid.

In one embodiment, the invention features the use of the compounds of the invention in a composition comprising surface-modified liposomes containing poly (ethylene glycol) lipids (PEG-modified, or long-circulating liposomes or stealth liposomes). In another embodiment, the invention features the use of compounds of the invention covalently attached to polyethylene glycol. These formulations offer a method for increasing the accumulation of drugs in target tissues. This class of drug carriers resists opsonization and elimination by the mononuclear phagocytic system (MPS or RES), thereby enabling longer blood circulation times and enhanced tissue exposure for the encapsulated drug (Lasic et al. Chem. Rev. 1995, 95, 2601-2627; Ishiwataet al., Chem. Pharm. Bull. 1995, 43, 1005-1011). Such compositions have been shown to accumulate selectively in tumors, presumably by extravasation and capture in the neovascularized target tissues (Lasic et al., Science 1995, 267, 1275-1276; Oku et al.,1995, Biochim. Biophys. Acta, 1238, 86-90). The long-circulating compositions enhance the pharmacokinetics and pharmacodynamics of therapeutic compounds, such as DNA and RNA, particularly compared to conventional cationic liposomes which are known to accumulate in tissues of the MPS (Liu et al., J. Biol. Chem. 1995, 42, 24864-24870; Choi et al., International PCT Publication No. WO 96/10391; Ansell et al., International PCT Publication No. WO 96/10390; Holland et al., International PCT Publication No. WO 96/10392). Long-circulating compositions are also likely to protect drugs from nuclease degradation to a greater extent compared to cationic liposomes, based on their ability to avoid accumulation in metabolically aggressive MPS tissues such as the liver and spleen.

The present invention also includes a composition(s) prepared for storage or administrationthat includes a pharmaceutically effective amount of the desired compound(s) in a pharmaceutically acceptable carrier or diluent. Acceptable carriers or diluents for therapeutic use are well known in the pharmaceutical art, and are described, for example, in Remington's Pharmaceutical Sciences, Mack Publishing Co. (A.R. Gennaro edit. 1985) hereby incorporated by reference herein. For example, preservatives, stabilizers, dyes and flavoring agents can be included in the composition. Examples of such agents include but are not limited to sodium benzoate, sorbic acid and esters of p-hydroxybenzoic acid. In addition, antioxidants and suspending agents can be included in the composition.

A pharmaceutically effective dose is that dose required to prevent, inhibit the occurrence, or treat (alleviate a symptom to some extent, preferably all of the symptoms) of a disease state. The pharmaceutically effective dose depends on the type of disease, the composition used, the route of administration, the type of mammal being treated, the

5   physical characteristics of the specific mammal under consideration, concurrent medication, and other factors which those skilled in the medical arts will recognize. Generally, an amount between 0.1 mg/kg and 100 mg/kg body weight/day of active ingredients is administered dependent upon potency of the negatively charged polymer. Furthermore, the compounds of the invention and formulations thereof can be

10   administered to a fetus via administration to the mother of a fetus.

The compounds of the invention and formulations thereof can be administered orally, topically, parenterally, by inhalation or spray or rectally in dosage unit formulations containing conventional non-toxic pharmaceutically acceptable carriers, adjuvants and vehicles. The term parenteral as used herein includes percutaneous,

15   subcutaneous, intravascular (e.g., intravenous), intramuscular, or intrathecal injection or infusion techniques and the like. In addition, there is provided a pharmaceutical formulation comprising a nucleic acid molecule of the invention and a pharmaceutically acceptable carrier. One or more nucleic acid molecules of the invention can be present in association with one or more non-toxic pharmaceutically acceptable carriers and/or

20   diluents and/or adjuvants, and if desired other active ingredients. The pharmaceutical compositions containing nucleic acid molecules of the invention can be in a form suitable for oral use, for example, as tablets, troches, lozenges, aqueous or oily suspensions, dispersible powders or granules, emulsion, hard or soft capsules, or syrups or elixirs.

Compositions intended for oral use can be prepared according to any method

25   known to the art for the manufacture of pharmaceutical compositions and such compositions can contain one or more such sweetening agents, flavoring agents, coloring agents or preservative agents in order to provide pharmaceutically elegant and palatable preparations. Tablets contain the active ingredient in admixture with non-toxic pharmaceutically acceptable excipients that are suitable for the manufacture of tablets.

30   These excipients can be, for example, inert diluents, such as calcium carbonate, sodium carbonate, lactose, calcium phosphate or sodium phosphate; granulating and disintegrating agents, for example, corn starch, or alginic acid; binding agents, for example starch, gelatin or acacia, and lubricating agents, for example magnesium stearate, stearic acid or talc. The tablets can be uncoated or they can be coated by known

35   techniques. In some cases such coatings can be prepared by known techniques to delay disintegration and absorption in the gastrointestinal tract and thereby provide a sustained

action over a longer period. For example, a time delay material such as glyceryl monosterate or glyceryl distearate can be employed.

Formulations for oral use can also be presented as hard gelatin capsules wherein the active ingredient is mixed with an inert solid diluent, for example, calcium carbonate, calcium phosphate or kaolin, or as soft gelatin capsules wherein the active ingredient is mixed with water or an oil medium, for example peanut oil, liquid paraffin or olive oil.

Aqueous suspensions contain the active materials in admixture with excipients suitable for the manufacture of aqueous suspensions. Such excipients are suspending agents, for example sodium carboxymethylcellulose, methylcellulose, hydropropyl-methylcellulose, sodium alginate, polyvinylpyrrolidone, gum tragacanth and gum acacia; dispersing or wetting agents can be a naturally-occurring phosphatide, for example, lecithin, or condensation products of an alkylene oxide with fatty acids, for example polyoxyethylene stearate, or condensation products of ethylene oxide with long chain aliphatic alcohols, for example heptadecaethyleneoxycetanol, or condensation products of ethylene oxide with partial esters derived from fatty acids and a hexitol such as polyoxyethylene sorbitol monooleate, or condensation products of ethylene oxide with partial esters derived from fatty acids and hexitol anhydrides, for example polyethylene sorbitan monooleate. The aqueous suspensions can also contain one or more preservatives, for example ethyl, or n-propyl p-hydroxybenzoate, one or more coloring agents, one or more flavoring agents, and one or more sweetening agents, such as sucrose or saccharin.

Oily suspensions can be formulated by suspending the active ingredients in a vegetable oil, for example arachis oil, olive oil, sesame oil or coconut oil, or in a mineral oil such as liquid paraffin. The oily suspensions can contain a thickening agent, for example beeswax, hard paraffin or cetyl alcohol. Sweetening agents and flavoring agents can be added to provide palatable oral preparations. These compositions can be preserved by the addition of an anti-oxidant such as ascorbic acid.

Dispersible powders and granules suitable for preparation of an aqueous suspension by the addition of water provide the active ingredient in admixture with a dispersing or wetting agent, suspending agent and one or more preservatives. Suitable dispersing or wetting agents or suspending agents are exemplified by those already mentioned above. Additional excipients, for example sweetening, flavoring and coloring agents, can also be present.

Pharmaceutical compositions of the invention can also be in the form of oil-in-water emulsions. The oily phase can be a vegetable oil or a mineral oil or mixtures of these. Suitable emulsifying agents can be naturally-occurring gums, for example gum acacia or gum tragacanth, naturally-occurring phosphatides, for example soy bean, lecithin, and esters or partial esters derived from fatty acids and hexitol, anhydrides, for example, sorbitan monooleate, and condensation products of the said partial esters with ethylene oxide, for example polyoxyethylene sorbitan monooleate. The emulsions can also contain sweetening and flavoring agents.

Syrups and elixirs can be formulated with sweetening agents, for example glycerol, propylene glycol, sorbitol, glucose or sucrose. Such formulations can also contain a demulcent, a preservative and flavoring and coloring agents. The pharmaceutical compositions can be in the form of a sterile injectable aqueous or oleaginous suspension. This suspension can be formulated according to the known art using those suitable dispersing or wetting agents and suspending agents that have been mentioned above. The sterile injectable preparation can also be a sterile injectable solution or suspension in a non-toxic parentally acceptable diluent or solvent, for example as a solution in 1,3-butanediol. Among the acceptable vehicles and solvents that can be employed are water, Ringer's solution and isotonic sodium chloride solution. In addition, sterile, fixed oils are conventionally employed as a solvent or suspending medium. For this purpose any bland fixed oil can be employed including synthetic mono-or diglycerides. In addition, fatty acids such as oleic acid find use in the preparation of injectables.

The compounds of the invention can also be administered in the form of suppositories, e.g., for rectal administration of the drug. These compositions can be prepared by mixing the drug with a suitable non-irritating excipient that is solid at ordinary temperatures but liquid at the rectal temperature and will therefore melt in the rectum to release the drug. Such materials include cocoa butter and polyethylene glycols.

Compounds of the invention can be administered parenterally in a sterile medium. The drug, depending on the vehicle and concentration used, can either be suspended or dissolved in the vehicle. Advantageously, adjuvants such as local anesthetics, preservatives and buffering agents can be dissolved in the vehicle.

Dosage levels of the order of from about 0.1 mg to about 140 mg per kilogram of body weight per day are useful in the treatment of the above-indicated conditions (about 0.5 mg to about 7 g per patient per day). The amount of active ingredient that can be combined with the carrier materials to produce a single dosage form will vary depending

upon the host treated and the particular mode of administration. Dosage unit forms will generally contain between from about 1 mg to about 500 mg of an active ingredient.

5   It will be understood, however, that the specific dose level for any particular patient will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, sex, diet, time of administration, route of administration, and rate of excretion, drug combination and the severity of the particular disease undergoing therapy.

For administration to non-human animals, the composition can also be added to the animal feed or drinking water. It can be convenient to formulate the animal feed and
10   drinking water compositions so that the animal takes in a therapeutically appropriate quantity of the composition along with its diet. It can also be convenient to present the composition as a premix for addition to the feed or drinking water.

The compounds of the present invention can also be administered to a patient in combination with other therapeutic compounds to increase the overall therapeutic effect.
15   The use of multiple compounds to treat an indication can increase the beneficial effects while reducing the presence of side effects.

Synthesis of Nucleic acid Molecules

Synthesis of nucleic acids greater than 100 nucleotides in length is difficult using automated methods, and the therapeutic cost of such molecules is prohibitive. In this
20   invention, small nucleic acid motifs ("small refers to nucleic acid motifs less than about 100 nucleotides in length, preferably less than about 80 nucleotides in length, and more preferably less than about 50 nucleotides in length; e.g., antisense oligonucleotides, hammerhead or the NCH ribozymes) are preferably used for exogenous delivery. The simple structure of these molecules increases the ability of the nucleic acid to invade
25   targeted regions of RNA structure. Exemplary molecules of the instant invention are chemically synthesized, and others can similarly be synthesized.

Oligonucleotides (eg; antisense GeneBlocs) are synthesized using protocols known in the art as described in Caruthers et al., 1992, Methods in Enzymology 211, 3-19, Thompson et al., International PCT Publication No. WO 99/54459, Wincott et al., 1995,
30   Nucleic Acids Res. 23, 2677-2684, Wincott et al., 1997, Methods Mol. Bio., 74, 59, Brennan et al., 1998, Biotechnol Bioeng., 61, 33-45, and Brennan, US patent No. 6,001,311. All of these references are incorporated herein by reference. The synthesis of oligonucleotides makes use of common nucleic acid protecting and coupling groups, such as dimethoxytrityl at the 5'-end, and phosphoramidites at the 3'-end. In a non-limiting
35   example, small scale syntheses are conducted on a 394 Applied Biosystems, Inc.

synthesizer using a 0.2 μmol scale protocol with a 2.5 min coupling step for 2'-O-methylated nucleotides and a 45 sec coupling step for 2'-deoxy nucleotides. **Table II** outlines the amounts and the contact times of the reagents used in the synthesis cycle. Alternatively, syntheses at the 0.2 μmol scale can be performed on a 96-well plate synthesizer, such as the instrument produced by Protogene (Palo Alto, CA) with minimal modification to the cycle. In a non-limiting example, a 33-fold excess (60 μL of 0.11 M = 6.6 μmol) of 2'-O-methyl phosphoramidite and a 105-fold excess of S-ethyl tetrazole (60 μL of 0.25 M = 15 μmol) can be used in each coupling cycle of 2'-O-methyl residues relative to polymer-bound 5'-hydroxyl. In a non-limiting example, a 22-fold excess (40 μL of 0.11 M = 4.4 μmol) of deoxy phosphoramidite and a 70-fold excess of S-ethyl tetrazole (40 μL of 0.25 M = 10 μmol) can be used in each coupling cycle of deoxy residues relative to polymer-bound 5'-hydroxyl. Average coupling yields on the 394 Applied Biosystems, Inc. synthesizer, determined by colorimetric quantitation of the trityl fractions, are typically 97.5-99%. Other oligonucleotide synthesis reagents for the 394 Applied Biosystems, Inc. synthesizer includebut are not limited to; detritylation solution is 3% TCA in methylene chloride (ABI); capping is performed with 16% *N*-methyl imidazole in THF (ABI) and 10% acetic anhydride/10% 2,6-lutidine in THF (ABI); and oxidation solution is 16.9 mM I_2, 49 mM pyridine, 9% water in THF (PERSEPTIVE™). Burdick & Jackson Synthesis Grade acetonitrile is used directly from the reagent bottle. S-Ethyltetrazole solution (0.25 M in acetonitrile) is made up from the solid obtained from American International Chemical, Inc. Alternately, for the introduction of phosphorothioate linkages, Beaucage reagent (3H-1,2-Benzodithiol-3-one 1,1-dioxide, 0.05 M in acetonitrile) is used.

Deprotection of the antisense oligonucleotides is performed as follows: the polymer-bound trityl-on oligoribonucleotide is transferred to a 4 mL glass screw top vial and suspended in a solution of 40% aq. methylamine (1 mL) at 65 °C for 10 min. After cooling to –20 °C, the supernatant is removed from the polymer support. The support is washed three times with 1.0 mL of EtOH:MeCN:H2O/3:1:1, vortexed and the supernatant is then added to the first supernatant. The combined supernatants, containing the oligoribonucleotide, are dried to a white powder. Standard drying or lyophilization methods known to those skilled in the art can be used.

The method of synthesis used for normal RNA including certain enzymatic nucleic acid molecules follows the procedure as described in Usman *et al.,* 1987, *J. Am. Chem. Soc.,* 109, 7845; Scaringe *et al.,* 1990, *Nucleic Acids Res.,* 18, 5433; and Wincott *et al.,* 1995, *Nucleic Acids Res.* 23, 2677-2684 Wincott *et al.,* 1997, *Methods Mol. Bio.,* 74, 59, and makes use of common nucleic acid protecting and coupling groups, such as dimethoxytrityl at the 5'-end, and phosphoramidites at the 3'-end. In a non-limiting

example, small scale syntheses are conducted on a 394 Applied Biosystems, Inc. synthesizer using a 0.2 µmol scale protocol with a 7.5 min coupling step for alkylsilyl protected nucleotides and a 2.5 min coupling step for 2'-O-methylated nucleotides. **Table II** outlines the amounts and the contact times of the reagents used in the synthesis cycle. Alternatively, syntheses at the 0.2 µmol scale can be done on a 96-well plate synthesizer, such as the instrument produced by Protogene (Palo Alto, CA) with minimal modification to the cycle. A 33-fold excess (60 µL of 0.11 M = 6.6 µmol) of 2'-O-methyl phosphoramidite and a 75-fold excess of S-ethyl tetrazole (60 µL of 0.25 M = 15 µmol) can be used in each coupling cycle of 2'-O-methyl residues relative to polymer-bound 5'-hydroxyl. A 66-fold excess (120 µL of 0.11 M = 13.2 µmol) of alkylsilyl (ribo) protected phosphoramidite and a 150-fold excess of S-ethyl tetrazole (120 µL of 0.25 M = 30 µmol) can be used in each coupling cycle of ribo residues relative to polymer-bound 5'-hydroxyl. Average coupling yields on the 394 Applied Biosystems, Inc. synthesizer, determined by colorimetric quantitation of the trityl fractions, are typically 97.5-99%. Other oligonucleotide synthesis reagents for the 394 Applied Biosystems, Inc. synthesizer include; detritylation solution is 3% TCA in methylene chloride (ABI); capping is performed with 16% N-methyl imidazole in THF (ABI) and 10% acetic anhydride/10% 2,6-lutidine in THF (ABI); oxidation solution is 16.9 mM I_2, 49 mM pyridine, 9% water in THF (PERSEPTIVE™). Burdick & Jackson Synthesis Grade acetonitrile is used directly from the reagent bottle. S-Ethyltetrazole solution (0.25 M in acetonitrile) is made up from the solid obtained from American International Chemical, Inc. Alternately, for the introduction of phosphorothioate linkages, Beaucage reagent (3H-1,2-Benzodithiol-3-one 1,1-dioxide0.05 M in acetonitrile) is used.

Deprotection of the RNA is performed using either a two-pot or one-pot protocol. For the two-pot protocol, the polymer-bound trityl-on oligoribonucleotide is transferred to a 4 mL glass screw top vial and suspended in a solution of 40% aq. methylamine (1 mL) at 65 °C for 10 min. After cooling to −20 °C, the supernatant is removed from the polymer support. The support is washed three times with 1.0 mL of EtOH:MeCN:H2O/3:1:1, vortexed and the supernatant is then added to the first supernatant. The combined supernatants, containing the oligoribonucleotide, are dried to a white powder. The base deprotected oligoribonucleotide is resuspended in anhydrous TEA/HF/NMP solution (300 µL of a solution of 1.5 mL N-methylpyrrolidinone, 750 µL TEA and 1 mL TEA•3HF to provide a 1.4 M HF concentration) and heated to 65 °C. After 1.5 h, the oligomer is quenched with 1.5 M NH_4HCO_3.

Alternatively, for the one-pot protocol, the polymer-bound trityl-on oligoribonucleotide is transferred to a 4 mL glass screw top vial and suspended in a solution of 33% ethanolic methylamine/DMSO: 1/1 (0.8 mL) at 65 °C for 15 min. The

vial is brought to r.t. TEA•3HF (0.1 mL) is added and the vial is heated at 65 °C for 15 min. The sample is cooled at –20 °C and then quenched with 1.5 M NH$_4$HCO$_3$.

For purification of the trityl-on oligomers, the quenched NH$_4$HCO$_3$ solution is loaded onto a C-18 containing cartridge that had been prewashed with acetonitrile

5   followed by 50 mM TEAA. After washing the loaded cartridge with water, the RNA is detritylated with 0.5% TFA for 13 min. The cartridge is then washed again with water, salt exchanged with 1 M NaCl and washed with water again. The oligonucleotide is then eluted with 30% acetonitrile.

Inactive hammerhead ribozymes or binding attenuated control ((BAC)

10   oligonucleotides) are synthesized by substituting a U for G$_5$ and a U for A$_{14}$ (numbering from Hertel, K. J., *et al.*, 1992, *Nucleic Acids Res.*, 20, 3252). Similarly, one or more nucleotide substitutions can be introduced in other enzymatic nucleic acid molecules to inactivate the molecule and such molecules can serve as a negative control.

The average stepwise coupling yields are typically >98% (Wincott *et al.*, 1995

15   *Nucleic Acids Res.* 23, 2677-2684). Those of ordinary skill in the art will recognize that the scale of synthesis can be adapted to be larger or smaller than the example described above including, but not limited to, 96 well format, with the ratio of chemicals used in the reaction being adjusted accordingly.

Alternatively, the nucleic acid molecules of the present invention can be synthesized

20   separately and joined together post-synthetically, for example by ligation (Moore *et al.*, 1992, *Science* 256, 9923; Draper *et al.*, International PCT publication No. WO 93/23569; Shabarova *et al.*, 1991, *Nucleic Acids Research* 19, 4247; Bellon *et al.*, 1997, *Nucleosides & Nucleotides*, 16, 951; Bellon *et al.*, 1997, *Bioconjugate Chem.* 8, 204).

The nucleic acid molecules of the present invention are modified extensively to

25   enhance stability by modification with nuclease resistant groups, for example, 2'-amino, 2'-*C*-allyl, 2'-flouro, 2'-*O*-methyl, 2'-H (for a review see Usman and Cedergren, 1992, *TIBS* 17, 34; Usman *et al.*, 1994, *Nucleic Acids Symp. Ser.* 31, 163). Ribozymes are purified by gel electrophoresis using general methods or are purified by high pressure liquid chromatography (HPLC; See Wincott *et al.*, Supra, the totality of which is hereby

30   incorporated herein by reference) and are re-suspended in water.

<u>Optimizing Activity of the nucleic acid molecule of the invention.</u>

Chemically synthesizing nucleic acid molecules with modifications (base, sugar and/or phosphate) that prevent their degradation by serum ribonucleases can increase their potency (see *e.g.,* Eckstein *et al.*, International Publication No. WO 92/07065; Perrault *et*

35   *al.*, 1990 *Nature* 344, 565; Pieken *et al.*, 1991, *Science* 253, 314; Usman and Cedergren, 1992, *Trends in Biochem. Sci.* 17, 334; Usman *et al.*, International Publication No.

WO 93/15187; and Rossi *et al.*, International Publication No. WO 91/03162; Sproat, US Patent No. 5,334,711; and Burgin *et al.*, *supra;* all of these describe various chemical modifications that can be made to the base, phosphate and/or sugar moieties of the nucleic acid molecules herein). Modifications which enhance their efficacy in cells, and removal

5   of bases from nucleic acid molecules to shorten oligonucleotide synthesis times and reduce chemical requirements are desired. (All these publications are hereby incorporated by reference herein).

There are several examples in the art describing sugar, base and phosphate modifications that can be introduced into nucleic acid molecules with significant

10   enhancement in their nuclease stability and efficacy. For example, oligonucleotides are modified to enhance stability and/or enhance biological activity by modification with nuclease resistant groups, for example, 2'-amino, 2'-*C*-allyl, 2'-flouro, 2'-*O*-methyl, 2'-H, nucleotide base modifications (for a review see Usman and Cedergren, 1992, *TIBS.* 17, 34; Usman *et al.*, 1994, *Nucleic Acids Symp. Ser.* 31, 163; Burgin *et al.*, 1996,

15   *Biochemistry* , 35, 14090). Sugar modification of nucleic acid molecules have been extensively described in the art (see Eckstein *et al.*, *International Publication* PCT No. WO 92/07065; Perrault *et al.* *Nature*, 1990, *344*, 565-568; Pieken *et al.* *Science*, 1991, *253*, 314-317; Usman and Cedergren, *Trends in Biochem. Sci.* , 1992, *17*, 334-339; Usman *et al. International Publication* PCT No. WO 93/15187; Sproat, *US Patent* No.

20   5,334,711 and Beigelman *et al.*, 1995, *J. Biol. Chem.*, 270, 25702; Beigelman *et al.*, International PCT publication No. WO 97/26270; Beigelman *et al.*, US Patent No. 5,716,824; Usman *et al.*, US patent No. 5,627,053; Woolf *et al.*, International PCT Publication No. WO 98/13526; Thompson *et al.*, USSN 60/082,404 which was filed on April 20, 1998; Karpeisky *et al.*, 1998, *Tetrahedron Lett.*, 39, 1131; Earnshaw and Gait,

25   1998, *Biopolymers (Nucleic acid Sciences),* 48, 39-55; Verma and Eckstein, 1998, *Annu. Rev. Biochem.*, 67, 99-134; and Burlina *et al.*, 1997, *Bioorg. Med. Chem.*, 5, 1999-2010; all of the references are hereby incorporated in their totality by reference herein). Such publications describe general methods and strategies to determine the location of incorporation of sugar, base and/or phosphate modifications and the like into ribozymes

30   without inhibiting catalysis, and are incorporated by reference herein. In view of such teachings, similar modifications can be used as described herein to modify the nucleic acid molecules of the instant invention.

While chemical modification of oligonucleotide internucleotide linkages with phosphorothioate, phosphorothioate, and/or 5'-methylphosphonate linkages improves

35   stability, too many of these modifications may cause some toxicity. Therefore, when designing nucleic acid molecules the amount of these internucleotide linkages should be

minimized. Without being bound by any particular theory, the reduction in the concentration of these linkages should lower toxicity resulting in increased efficacy and higher specificity of these molecules.

5   Nucleic acid molecules having chemical modifications that maintain or enhance activity are provided. Such nucleic acid is also generally more resistant to nucleases than unmodified nucleic acid. Thus, in a cell and/or *in vivo* the activity can not be significantly lowered. Therapeutic nucleic acid molecules (*e.g.*, enzymatic nucleic acid molecules and antisense nucleic acid molecules) delivered exogenously are optimally stable within cells until translation of the target RNA has been inhibited long enough to reduce the levels of

10   the undesirable protein. This period of time varies between hours to days depending upon the disease state. The nucleic acid molecules should be resistant to nucleases in order to function as effective intracellular therapeutic agents. Improvements in the chemical synthesis of RNA and DNA (Wincott *et al.*, 1995 *Nucleic Acids Res.* 23, 2677; Caruthers *et al.*, 1992, *Methods in Enzymology* 211,3-19 (incorporated by reference herein) have

15   expanded the ability to modify nucleic acid molecules by introducing nucleotide modifications to enhance their nuclease stability as described above.

Use of the nucleic acid-based molecules of the invention can lead to better treatment of the disease progression by affording the possibility of combination therapies (*e.g.*, multiple antisense or enzymatic nucleic acid molecules targeted to different genes,

20   nucleic acid molecules coupled with known small molecule inhibitors, or intermittent treatment with combinations of molecules (including different motifs) and/or other chemical or biological molecules). The treatment of patients with nucleic acid molecules can also include combinations of different types of nucleic acid molecules.

In another embodiment, nucleic acid catalysts having chemical modifications that

25   maintain or enhance enzymatic activity are provided. Such nucleic acids are also generally more resistant to nucleases than unmodified nucleic acid. Thus, in a cell and/or *in vivo* the activity of the nucleic acid can not be significantly lowered. As exemplified herein such enzymatic nucleic acids are useful in a cell and/or *in vivo* even if activity over all is reduced 10 fold (Burgin *et al.*, 1996, *Biochemistry*, 35, 14090). Such enzymatic

30   nucleic acids herein are said to "maintain" the enzymatic activity of an all RNA ribozyme or all DNA DNAzyme.

In another aspect the nucleic acid molecules comprise a 5' and/or a 3'- cap structure.

In another embodiment the 3'-cap includes, for example 4',5'-methylene nucleotide; 1-(beta-D-erythrofuranosyl) nucleotide; 4'-thio nucleotide, carbocyclic nucleotide; 5'-amino-alkyl phosphate; 1,3-diamino-2-propyl phosphate, 3-aminopropyl phosphate; 6-aminohexyl phosphate; 1,2-aminododecyl phosphate; hydroxypropyl phosphate; 1,5-anhydrohexitol nucleotide; L-nucleotide; alpha-nucleotide; modified base nucleotide; phosphorodithioate; *threo*-pentofuranosyl nucleotide; acyclic 3',4'-seco nucleotide; 3,4-dihydroxybutyl nucleotide; 3,5-dihydroxypentyl nucleotide, 5'-5'-inverted nucleotide moiety; 5'-5'-inverted abasic moiety; 5'-phosphoramidate; 5'-phosphorothioate; 1,4-butanediol phosphate; 5'-amino; bridging and/or non-bridging 5'-phosphoramidate, phosphorothioate and/or phosphorodithioate, bridging or non bridging methylphosphonate and 5'-mercapto moieties (for more details see Beaucage and Iyer, 1993, *Tetrahedron* 49, 1925; incorporated by reference herein).

In one embodiment, the invention features modified enzymatic nucleic acid molecules with phosphate backbone modifications comprising one or more phosphorothioate, phosphorodithioate, methylphosphonate, morpholino, amidate carbamate, carboxymethyl, acetamidate, polyamide, sulfonate, sulfonamide, sulfamate, formacetal, thioformacetal, and/or alkylsilyl, substitutions. For a review of oligonucleotide backbone modifications see Hunziker and Leumann, 1995, *Nucleic Acid Analogues: Synthesis and Properties,* in *Modern Synthetic Methods,* VCH, 331-417, and Mesmaeker *et al.,* 1994, *Novel Backbone Replacements for Oligonucleotides,* in *Carbohydrate Modifications in Antisense Research,* ACS, 24-39. These references are hereby incorporated by reference herein.

In connection with 2'-modified nucleotides as described for the invention, by "amino" is meant 2'-NH$_2$ or 2'-O- NH$_2$, which can be modified or unmodified. Such modified groups are described, for example, in Eckstein *et al.,* U.S. Patent 5,672,695 and Matulic-Adamic *et al.,* WO 98/28317, respectively, which are both incorporated by reference in their entireties.

Various modifications to nucleic acid (*e.g.,* antisense and ribozyme) structure can be made to enhance the utility of these molecules. For example, such modifications can enhance shelf-life, half-life *in vitro,* stability, and ease of introduction of such oligonucleotides to the target site, including *e.g.,* enhancing penetration of cellular membranes and conferring the ability to recognize and bind to targeted cells.

Use of these molecules can lead to better treatment of disease progression by affording the possibility of combination therapies (e.g., multiple enzymatic nucleic acid molecules targeted to different genes, enzymatic nucleic acid molecules coupled with

known small molecule inhibitors, or intermittent treatment with combinations of enzymatic nucleic acid molecules (including different enzymatic nucleic acid molecule motifs) and/or other chemical or biological molecules). The treatment of patients with nucleic acid molecules can also include combinations of different types of nucleic acid

5   molecules. Therapies can be devised which include a mixture of enzymatic nucleic acid molecules (including different enzymatic nucleic acid molecule motifs), antisense and/or 2-5A chimera molecules to one or more targets to alleviate symptoms of a disease.

Indications

10   Particular disease states that can be treated using compounds and compositions of the invention include, but are not limited to, cancers and cancerous conditions such as breast, lung, prostate, colorectal, brain, esophageal, stomach, bladder, pancreatic, cervical, head and neck, and ovarian cancer, melanoma, lymphoma, glioma, multidrug resistant cancers, and/or viral infections including HIV, HBV, HCV, CMV, RSV, HSV, poliovirus,

15   influenza, rhinovirus, west nile virus, Ebola virus, foot and mouth virus, and papilloma virus infection.

The molecules of the invention can be used in conjunction with other known methods, therapies, or drugs. For example, the use of monoclonal antibodies (eg; mAb

20   IMC C225, mAB ABX-EGF) treatment, tyrosine kinase inhibitors (TKIs), for example OSI-774 and ZD1839, chemotherapy, and/or radiation therapy, are all non-limiting examples of a methods that can be combined with or used in conjunction with the compounds of the instant invention. Common chemotherapies that can be combined with nucleic acid molecules of the instant invention include various combinations of cytotoxic

25   drugs to kill the cancer cells. These drugs include, but are not limited to, paclitaxel (Taxol), docetaxel, cisplatin, methotrexate, cyclophosphamide, doxorubin, fluorouracil carboplatin, edatrexate, gemcitabine, vinorelbine etc. Those skilled in the art will recognize that other drug compounds and therapies can be similarly be readily combined with the compounds of the instant invention are hence within the scope of the instant

30   invention.

Diagnostic uses

The compounds of this invention, for example, nucleic acid conjugate molecules, can be used as diagnostic tools to examine genetic drift and mutations within diseased

35   cells or to detect the presence of a disease related RNA in a cell. The close relationship between, for example, enzymatic nucleic acid molecule activity and the structure of the

target RNA allows the detection of mutations in any region of the molecule which alters the base-pairing and three-dimensional structure of the target RNA. By using multiple enzymatic nucleic acid molecules conjugates of the invention, one can map nucleotide changes which are important to RNA structure and function in vitro, as well as in cells

5 and tissues. Cleavage of target RNAs with enzymatic nucleic acid molecules can be used to inhibit gene expression and define the role (essentially) of specified gene products in the progression of disease. In this manner, other genetic targets can be defined as important mediators of the disease. These experiments can lead to better treatment of the disease progression by affording the possibility of combinational therapies (e.g., multiple

10 enzymatic nucleic acid molecules targeted to different genes, enzymatic nucleic acid molecules coupled with known small molecule inhibitors, or intermittent treatment with combinations of enzymatic nucleic acid molecules and/or other chemical or biological molecules). Other in vitro uses of enzymatic nucleic acid molecules of this invention are well known in the art, and include detection of the presence of mRNAs associated with a

15 disease-related condition. Such RNA is detected by determining the presence of a cleavage product after treatment with an enzymatic nucleic acid molecule using standard methodology.

In a specific example, enzymatic nucleic acid molecules that are delivered to cells

20 as conjugates and which cleave only wild-type or mutant forms of the target RNA are used for the assay. The first enzymatic nucleic acid molecule is used to identify wild-type RNA present in the sample and the second enzymatic nucleic acid molecule is used to identify mutant RNA in the sample. As reaction controls, synthetic substrates of both wild-type and mutant RNA are cleaved by both enzymatic nucleic acid molecules to

25 demonstrate the relative enzymatic nucleic acid molecule efficiencies in the reactions and the absence of cleavage of the "non-targeted" RNA species. The cleavage products from the synthetic substrates also serve to generate size markers for the analysis of wild-type and mutant RNAs in the sample population. Thus each analysis requires two enzymatic nucleic acid molecules, two substrates and one unknown sample which is combined into

30 six reactions. The presence of cleavage products is determined using an RNAse protection assay so that full-length and cleavage fragments of each RNA can be analyzed in one lane of a polyacrylamide gel. It is not absolutely required to quantify the results to gain insight into the expression of mutant RNAs and putative risk of the desired phenotypic changes in target cells. The expression of mRNA whose protein product is

35 implicated in the development of the phenotype is adequate to establish risk. If probes of comparable specific activity are used for both transcripts, then a qualitative comparison of RNA levels will be adequate and will decrease the cost of the initial diagnosis. Higher

mutant form to wild-type ratios are correlated with higher risk whether RNA levels are compared qualitatively or quantitatively. The use of enzymatic nucleic acid molecules in diagnostic applications contemplated by the instant invention is more fully described in George et al., US Patent Nos. 5,834,186 and 5,741,679, Shih et al., US Patent No. 5,589,332, Nathan et al., US Patent No 5,871,914, Nathan and Ellington, International PCT publication No. WO 00/24931, Breaker et al., International PCT Publication Nos. WO 00/26226 and 98/27104, and Sullenger et al., International PCT publication No. WO 99/29842.

Additional Uses

Potential uses of sequence-specific enzymatic nucleic acid molecules of the instant invention that are delivered to cells as conjugates can have many of the same applications for the study of RNA that DNA restriction endonucleases have for the study of DNA (Nathans et al., 1975 Ann. Rev. Biochem. 44:273). For example, the pattern of restriction fragments can be used to establish sequence relationships between two related RNAs, and large RNAs can be specifically cleaved to fragments of a size more useful for study. The ability to engineer sequence specificity of the enzymatic nucleic acid molecule is ideal for cleavage of RNAs of unknown sequence. Applicant has described the use of nucleic acid molecules to down-regulate gene expression of target genes in bacterial, microbial, fungal, viral, and eukaryotic systems including plant, or mammalian cells.

Example 1: Synthesis of O^1-(4-monomethoxytrityl)-N-(6-(N-(α-OFm-L-glutamyl)aminocaproyl))-D-threoninol-N^2-iBu-N^{10}-TFA-pteroic acid conjugate 3'-O-(2-cyanoethyl-N,N-diisopropylphosphor-amidite) (20) (Figure 5)

General. All reactions were carried out under a positive pressure of argon in anhydrous solvents. Commercially available reagents and anhydrous solvents were used without further purification. ^{1}H (400.035 MHz) and ^{31}P (161.947 MHz) NMR spectra were recorded in CDCl$_3$, unless stated otherwise, and chemical shifts in ppm refer to TMS and H$_3$PO$_4$, respectively. Analytical thin-layer chromatography (TLC) was performed with Merck Art.5554 Kieselgel 60 F$_{254}$ plates and flash column chromatography using Merck 0.040-0.063 mm silica gel 60.

N-(N-Fmoc-6-aminocaproyl)-D-threoninol (13). N-Fmoc-6-aminocaproic acid (10 g, 28.30 mmol) was dissolved in DMF (50 ml) and N-hydroxysuccinimide (3.26 g, 28.30 mmol) and 1,3-dicyclohexylcarbodiimide (5.84 g, 28.3 mmol) were added to the solution. The reaction mixture was stirred at RT (about 23°C) overnight and the

precipitated 1,3-dicyclohexylurea filtered off. To the filtrate D-threoninol (2.98 g, 28.30 mmol) was added and the reaction mixture stirred at RT overnight. The solution was reduced to *ca* half the volume *in vacuo*, the residue diluted with about m ml of ethyl acetate and extracted with about x ml of 5% NaHCO$_3$, followed by washing with brine. The organic layer was dried (Na$_2$SO$_4$), evaporated to a syrup and chromatographed by silica gel column chromatography using 1-10% gradient of methanol in ethyl acetate. Fractions containing the product were pooled and evaporated to a white solid (9.94 g, 80%). ^{1}H-NMR (DMSO-d$_6$-D$_2$O) δ7.97-7.30 (m, 8H, aromatic), 4.34 (d, J=6.80, 2H, Fm), 4.26 (t, J=6.80, 1H, Fm), 3.9 (m, 1H, H3 Thr), 3.69 (m, 1H, H2 Thr), 3.49 (dd, J=10.6, J=7.0, 1H, H1 Thr), 3.35 (dd, J=10.6, J=6.2, 1H, H1' Thr), 3.01 (m, 2H, *CH*$_2$CO Acp), 2.17 (m, 2H, *CH*$_2$NH Acp), 1.54 (m, 2H, CH$_2$ Acp), 1.45 (m, 2H, CH$_2$ Acp), 1.27 (m, 2H, CH$_2$ Acp), 1.04 (d, J=6.4, 3H, CH$_3$). MS/ESI$^+$ m/z 441.0 (M+H)$^+$.

O^1-(4-Monomethoxytrityl)-*N*-(*N*-Fmoc-6-aminocaproyl)-D-threoninol (14). To the solution of 13 (6 g, 13.62 mmol) in dry pyridine (80 ml) *p*-anisylchlorodiphenyl-methane (6 g, 19.43 mmol) was added and the reaction mixture stirred at RT overnight. Methanol was added (20 ml) and the solution concentrated *in vacuo*. The residual syrup was partitioned between about x ml of dichloromethane and about x ml of 5% NaHCO$_3$, the organic layer was washed with brine, dried (Na$_2$SO$_4$) and evaporated to dryness. Flash column chromatography using 1-3% gradient of methanol in dichloromethane afforded 14 as a white foam (6 g, 62%). ^{1}H-NMR (DMSO) δ7.97-6.94 (m, 22H, aromatic), 4.58 (d, 1H, J=5.2, OH), 4.35 (d, J=6.8, 2H, Fm), 4.27 (t, J=6.8, 1H, Fm), 3.97 (m, 2H, H2,H3 Thr), 3.80 (s, 3H, OCH$_3$), 3.13 (dd, J=8.4, J=5.6, 1H, H1 Thr), 3.01 (m, 2H, *CH*$_2$CO Acp), 2.92 (m, dd, J=8.4, J=6.4, 1H, H1' Thr), 2.21 (m, 2H, *CH*$_2$NH Acp), 1.57 (m, 2H, CH$_2$ Acp), 1.46 (m, 2H, CH$_2$ Acp), 1.30 (m, 2H, CH$_2$ Acp), 1.02 (d, J=5.6, 3H, CH$_3$). MS/ESI$^+$ m/z 735.5 (M+Na)$^+$.

O^1-(4-Monomethoxytrityl)-*N*-(6-aminocaproyl)-D-threoninol (15). 14 (9.1 g, 12.77 mmol) was dissolved in DMF (100 ml) containing piperidine (10 ml) and the reaction mixture was kept at RT for about 1 hour. The solvents were removed *in vacuo* and the residue purified by silica gel column chromatography using 1-10% gradient of methanol in dichloromethane to afford 15 as a syrup (4.46 g, 71%). ^{1}H-NMR δ7.48-6.92 (m, 14H, aromatic), 6.16 (d, J=8.8, 1H, NH), 4.17 (m, 1H, H3 Thr), 4.02 (m, 1H, H2 Thr), 3.86 (s, 3H, OCH$_3$), 3.50 (dd, J=9.7, J=4.4, 1H, H1 Thr), 3.37 (dd, J=9.7, J=3.4, 1H, H1' Thr), 2.78 (t, J=6.8, 2H, *CH*$_2$CO Acp), 2.33 (t, J=7.6, 2H, *CH*$_2$NH Acp), 1.76 (m, 2H, CH$_2$ Acp), 1.56 (m, 2H, CH$_2$ Acp), 1.50 (m, 2H, CH$_2$ Acp), 1.21 (d, J=6.4, 3H, CH$_3$). MS/ESI$^+$ m/z 491.5 (M+H)$^+$.

*O*¹-(4-Monomethoxytrityl)-*N*-(6-(*N*-(*N*-Boc-α-OFm-L-glutamyl)
aminocaproyl))-D-threoninol (16). To the solution of *N*-Boc-α-OFm-glutamic acid
(Bachem) (1.91 g, 4.48 mmol) in DMF (10 ml) *N*-hydroxysuccinimide (518 mg, 4.50
mmol) and 1,3-dicyclohexylcarbodiimide (928 mg, 4.50 mmol) was added and the
reaction mixture was stirred at RT overnight. 1,3-Dicyclohexylurea was filtered off and to
the filtrate 15 (2 g, 4.08 mmol) and pyridine (2 ml) were added. The reaction mixture was
stirred at RT for 3 hours and than concentrated *in vacuo*. The residue was partitioned
between ethyl acetate and 5% Na₂HCO₃, the organic layer extracted with brine as
previously described, dried (Na₂SO₄) and evaporated to a syrup. Column chromatography
using 2-10% gradient of methanol in dichlotomethane afforded 16 as a white foam (3.4 g,
93%). ¹H-NMR δ 7.86-6.91 (m, 22H, aromatic), 6.13 (d, J=8.8, 1H, NH), 5.93 (br s, 1H,
NH), 5.43 (d, J=8.4, 1H, NH), 4.63 (dd, J=10.6, J=6.4, 1H, Fm), 4.54 (dd, J=10.6, J=6.4,
1H, Fm), 4.38 (m, 1H, Glu), 4.3 (t, J=6.4, 1H, Fm), 4.18 (m, 1H, H3 Thr), 4.01 (m, 1H,
H2 Thr), 3.88 (s, 3H, OCH₃), 3.49 (dd, J=9.5, J=4.4, 1H, H1 Thr), 3.37 (dd, J=9.5, J=3.8,
1H, H1' Thr), 3.32 (m, 2H, *CH₂*CO Acp), 3.09 (br s, 1H, OH), 2.32 (m, 2H, *CH₂*NH
Acp), 2.17 (m, 3H, Glu), 1.97 (m, 1H, Glu), 1.77 (m, 2H, CH₂ Acp), 1.61 (m, 2H, CH₂
Acp), 1.52 (s, 9H, t-Bu), 1.21 (d, J=6.4, 3H, CH₃). MS/ESI⁺ m/z 920.5 (M+Na)⁺.

N-(6-(*N*-α-OFm-L-glutamyl)aminocaproyl))-D-threoninol hydrochloride (17).
16 (2 g, 2.23 mmol) was dissolved in methanol (30 ml) containing anisole (10 ml) and to
this solution x ml of 4M HCl in dioxane was added. The reaction mixture was stirred for
3 hours at RT and then concentrated *in vacuo*. The residue was dissolved in ethanol and
the product precipitated by addition of x ml of ether. The precipitate was washed with
ether and dried to give 17 as a colorless foam (1 g, 80%). ¹H-NMR (DMSO-d₆-D₂O)
δ7.97-7.40 (m, 8H, aromatic), 4.70 (m, 1H, Fm), 4.55 (m, 1H, Fm), 4.40 (t, J=6.4, 1H,
Fm), 4.14 (t, J=6.6, 1H, Glu), 3.90 (dd, J=2.8, J=6.4, 1H, H3 Thr), 3.68 (m, 1H, H2 Thr),
3.49 (dd, J=10.6, J=7.0, 1H, H1 Thr), 3.36 (dd, J=10.6, J=6.2, 1H, H1' Thr), 3.07 (m, 2H,
*CH₂*CO Acp), 2.17 m, 3H), 1.93 (m, 2H), 1.45 (m, 2H), 1.27 (m, 2H), 1.04 (d, J=6.4, 3H
Thr). MS/ESI⁺ m/z 526.5 (M+H)⁺.

N-(6-(*N*-α-OFm-L-glutamyl)aminocaproyl))-D-threoninol-*N*²-*i*Bu-*N*¹⁰-TFA-
pteroic acid conjugate (18). To the solution of *N*²-*i*Bu-*N*¹⁰-TFA-pteroic acid[1] (480 mg, 1
mmol) in DMF (5 ml) 1-hydroxybenzotriazole (203 mg, 1.50 mmol), EDCI (288 mg, 1.50
mmol) and 17 (free base, 631 mg, 1.2 mmol) are added. The reaction mixture is stirred at
RT for 2 hours, then concentrated to *ca* 3 ml and loaded on the column of silica gel.
Elution with dichloromethane, followed by 1-20% gradient of methanol in
dichloromethane afforded 18 (0.5 g, 51%). ¹H-NMR (DMSO-d₆-D₂O) δ 9.09 (d, J=6.8,
1H, NH) 8.96 (s, 1H, H7 pteroic acid), 8.02-7.19 (m, 13H, aromatic, NH), 5.30 (s, 2H,

pteroic acid), 4.50 (m, 1H, Glu), 4.41 (d, J=6.8, 2H, Fm), 4.29 (t, J=6.8, 1H, Fm), 3.89 (dd, J=6.2, J=2.8, 1H, H3 Thr), 3.68 (m, 1H, H2 Thr), 3.48 (dd, J=10.4, J=7.0, 1H, H1 Thr), 3.36 (dd, J=10.4, J=6.2, 1H H1' Thr), 3.06 (m, 2H, CH_2CO Acp), 2.84 (m, 1H, iBu), 2.25 (m, 2H, CH_2NH Acp), 2.16 (m, 3H, Glu), 1.99 (m, 1H, Glu), 1.52 (m, 2H Acp), 1.42 (m, 2H Acp), 1.27 (m, 2H Acp), 1.20 (s, 3H iBu), 1.19 (s, 3H, iBu), 1.03 (d, J=6.2, 3H Thr). MS/ESI⁻ m/z 984.5 (M-H)⁻.

O^1-(4-monomethoxytrityl)-N-(6-(N-α-OFm-L-glutamyl)aminocaproyl))-D-threoninol-N^2-iBu-N^{10}-TFA-pteroic acid conjugate (19). To the solution of conjugate 18 (1 g, 1.01 mmol) in dry pyridine (15 ml) p-anisylchlorodiphenylmethane (405 mg) was added and the reaction mixture was stirred, protected from moisture, at RT overnight. Methanol (3 ml) was added and the reaction mixture concentrated to a syrup *in vacuo*. The residue was partitioned between dichloromethane and 5% NaHCO₃, the organic layer washed with brine, dried (Na₂SO₄) and evaporated to dryness. Column chromatography using 0.5-10% gradient of methanol in dichloromethane afforded 19 as a colorless foam (0.5 g, 39%. ^{1}H-NMR (DMSO-d₆-D₂O δ9.09 (d, J=6.8, 1H, NH) 8.94 (s, 1H, H7 pteroic acid), 8.00-6.93 (m, 27H, aromatic, NH), 5.30 (s, 2H, pteroic acid), 4.50 (m, 1H, Glu), 4.40 (d, J=6.8, 2H, Fm), 4.29 (t, J=6.8, 1H, Fm), 3.94 (m, 2H, H3,H2 Thr), 3.79 (s, 3H, OCH₃) 3.11 (dd, J=8.6, J=5.8, 1H, H1 Thr), 3.04 (m, 2H, CH_2CO Acp), 2.91 (dd, J=8.6, J=6.4, 1H, H1' Thr), 2.85 (m, 1H, iBu), 2.25 (m, 2H, CH_2NH Acp), 2.19 (m, 2H, Glu), 2.13 (m, 1H, Glu), 1.98 (m, 1H, Glu), 1.55 (m, 2H Acp), 1.42 (m, 2H Acp), 1.29(m, 2H Acp), 1.20 (s, 3H iBu), 1.18 (s, 3H, iBu), 1.00 (d, J=6.4, 3H Thr). MS/ESI⁻ m/z 1257.0 (M-H)⁻.

O^1-(4-monomethoxytrityl)-N-(6-(N-α-OFm-L-glutamyl)aminocaproyl))-D-threoninol-N^2-iBu-N^{10}-TFA-pteroic acid conjugate 3'-O-(2-cyanoethyl-N,N-diisopropylphosphor-amidite) (20). To the solution of 19 (500 mg, 0.40 mmol) in dichloromethane (2 ml) 2-cyanoethyl tetraisopropylphosphordiamidite (152 μL, 0.48 mmol) was added followed by pyridinium trifluoroacetate (93 mg, 0.48 mmol). The reaction mixture was stirred at RT for 1 hour and than loaded on the column of silica gel in hexanes. Elution using ethyl acetate-hexanes 1:1, followed by ethyl acetate and ethyl acetate-acetone 1:1 in the presence of 1% pyridine afforded 20 as a colorless foam (480 mg, 83%). ^{31}P NMR δ 149.4 (s), 149.0 (s).

Example 2: Synthesis of 2-dithiopyridyl activated folic acid (30) (Figure 9)

Synthesis of the cysteamine modified folate 30 is presented in Fig. 9. Monomethoxytrityl cysteamine 21 was prepared by selective tritylation of the thiol group of cysteamine with 4-methoxytrityl alcohol in trifluoroacetic acid. Peptide coupling of 21

with Fmoc-Glu-O*t*Bu (Bachem Bioscience Inc., King of Prussia, PA) in the presence of PyBOP yielded **22** in a high yield. *N*-Fmoc group was removed smoothly with piperidine to give **23**. Condensation of **23** with *p*-(4-methoxytrityl)aminobenzoic acid, prepared by reaction of *p*-aminobenzoic acid with 4-methoxytrityl chloride in pyridine, afforded the

5   fully protected conjugate **24**. Selective cleavage of *N*-MMTr group with acetic acid afforded **25** in quantitative yield. Shiff base formation between **25** and N^2-*i*Bu-6-formylpterin **26**,[9] followed by reduction with borane-pyridine complex proceeded with a good yield to give fully protected cysteamine-folate adduct **27**.[12] The consecutive cleavage of protecting groups of **27** with base and acid yielded thiol derivative **29**. The

10 thiol exchange reaction of **29** with 2,2-dipyridyl disulfide afforded the desired *S*-pyridyl activated synthon **30** as a yellow powder; Isolated as a TEA$^+$ salt: ^{1}H NMR spectrum for **10** in D$_2$O: δ 8.68 (s, 1 H, H-7), 8.10 (d, J = 3.6, 1 H, pyr), 7.61 (d, J = 8.8, 2 H, PABA), 7.43 (m, 1 H, pyr), 7.04 (d, J = 7.6, 1 H, pyr), 6.93 (m, 1 H, pyr), 6.82 (d, J = 8.8, 1 H, PABA), 4.60 (s, 2 H, 6-CH$_2$), 4.28 (m, 1 H, Glu), 3.30-3.08 (m, 2 H, cysteamine), 3.05 (m,

15 6 H, TEA), 2.37 (m, 2 H, cysteamine), 2.10 (m, 4 H, Glu), 1.20 (m, 9 H, TEA). MS/ESI$^-$ m/z 608.02 [M-H]$^-$. It is worth noting that the isolation of **30** as its TEA$^+$ or Na$^+$ salt made it soluble in DMSO and/or water, which is an important requirement for its use in conjugation reactions.

Example 3: Post synthetic conjugation of enzymatic nucleic acid to form nucleic acid-
20 ### folate conjugate (33) (**Figure 10**)

Oligonucleotide synthesis, deprotection and purification was performed as described herein. 5'-Thiol-Modifier C6 (Glen Research, Sterling, Virginia) was coupled as the last phosphoramidite to the 5'-end of a growing oligonucleotide chain. After cleavage from the solid support and base deprotection, the disulfide modified enzymatic

25 nucleic acid molecule **31** (Fig. 10) was purified using ion exchange chromatography. The thiol group was unmasked by reduction with dithiothreitol (DTT) to afford **32** which was purified by gel filtration and immediately conjugated with **30**. The resulting conjugate **33** was separated from the excess folate by gel filtration and then purified by RP HPLC using gradient of acetonitrile in 50 mM triethylammonium acetate (TEAA). Desalting was

30 performed by RP HPLC. Reactions were conducted on 400 mg of disulfide modified enzymatic nucleic acid molecule **31** to afford 200-250 mg (50-60% yield) of conjugate **33**. MALDI TOF MS confirmed the structure: **13** [M-H]$^-$ 12084.74 (calc.12083.82). An alternative approach to this synthesis is shown in **Figure 11**.

As shown in Examples 2 and 3, a folate-cysteamine adduct can be prepared by a
35 scaleable solution phase synthesis in a good overall yield. Disulfide conjugation of this

novel targeting ligand to the thiol-modified oligonucleotide is suitable for the multi-gram scale synthesis. The 9-atom spacer provides a useful spatial separation between folate and attached oligonucleotide cargo. Importantly, conjugation of folate to the oligonucleotide through a disulfide bond should permit intermolecular separation which was suggested to

5   be required for the functional cytosolic entry of a protein drug.

Example 4: Synthesis of Galactose and N-acetyl-Galactosamine conjugates (**Figures 13, 14, and 15**)

Applicant has designed both nucleoside and non-nucleoside-*N*-acetyl-D-galactosamine conjugates suitable for incorporation at any desired position of an

10   oligonucleotide. Multiple incorporations of these monomers could result in a "glycoside cluster effect".

All reactions were carried out under a positive pressure of argon in anhydrous solvents. Commercially available reagents and anhydrous solvents were used without further purification. *N*-acetyl-D-galactosamine was purchased from Pfanstiel (Waukegan,

15   IL), folic acid from Sigma (St. Louis, MO), D-threoninol from Aldrich (Milwaukee, WI) and *N*-Boc-α-OFm glutamic acid from Bachem. ^{1}H (400.035 MHz) and ^{31}P (161.947 MHz) NMR spectra were recorded in $CDCl_3$, unless stated otherwise, and chemical shifts in ppm refer to TMS and H3PO4, respectively. Analytical thin-layer chromatography (TLC) was performed with Merck Art.5554 Kieselgel 60 F_{254} plates and flash column

20   chromatography using Merck 0.040-0.063 mm silica gel 60. The general procedures for RNA synthesis, deprotection and purification are described herein. MALDI-TOF mass spectra were determined on PerSeptive Biosystems Voyager spectrometer. Electrospray mass spectrometry was run on the PE/Sciex API365 instrument.

2'-(*N*-L-lysyl)amino-5'-*O*-4,4'-dimethoxytrityl-2'-deoxyuridine   (2).   2'-(*N*-α,ε-bis-

25   Fmoc-L-lysyl)amino-5'-*O*-4,4'-dimethoxytrityl-2'-deoxyuridine (**1**) (4 g, 3.58 mmol) was dissolved in anhydrous DMF (30 ml) and diethylamine (4 ml) was added. The reaction mixture was stirred at rt for 5 hours and than concentrated (oil pump) to a syrup. The residue was dissolved in ethanol and ether was added to precipitate the product (1.8 g, 75%). ^{1}H-NMR (DMSO-d6-D2O) δ 7.70 (d, $J_{6,5}$=8.4, 1H, H6), 7.48-6.95 (m, 13H,

30   aromatic), 5.93 (d, $J_{1',2'}$=8.4, 1H, H1'), 5.41 (d, $J_{5,6}$=8.4, 1H, H5), 4,62 (m, 1H, H2'), 4.19 (d, 1H, $J_{3',2'}$=6.0, H3'), 3.81 (s, 6H, 2×OMe), 3.30 (m, 4H, 2H5', CH_2), 1.60-1.20 (m, 6H, 3×CH_2). MS/ESI$^+$ m/z 674.0 (M+H)$^+$.

N-Acetyl-1,4,6-tri-*O*-acetyl-2-amino-2-deoxy-β-D-galactospyranose (3). *N*-Acetyl-D-galac-tosamine (6.77 g, 30.60 mmol) was suspended in acetonitrile (200 ml) and triethylamine (50 ml, 359 mmol) was added. The mixture was cooled in an ice-bath and

35

acetic anhydride (50 ml, 530 mmol)) was added dropwise under cooling. The suspension slowly cleared and was then stirred at rt for 2 hours. It was than cooled in an ice-bath and methanol (60 ml) was added and the stirring continued for 15 min. The mixture was concentrated under reduced pressure and the residue partitioned between dichloromethane

5   and 1 N HCl. Organic layer was washed twice with 5% NaHCO$_3$, followed by brine, dried (Na$_2$SO$_4$) and evaporated to dryness to afford 10 g (84%) of **3** as a colorless foam. ^{1}H NMR was in agreement with published data (Findeis, 1994, *Int. J. Peptide Protein Res.*, 43, 477-485.

2-Acetamido-3,4,6-tetra-*O*-acetyl-1-chloro-D-galactospyranose (4). This compound

10   was prepared from **3** as described by Findeis *supra*.

Benzyl 12-Hydroxydodecanoate (5). To a cooled (0 °C) and stirred solution of 12-hydroxydodecanoic acid (10.65 g, 49.2 mmol) in DMF (70 ml) DBU (8.2 ml, 54.1 mmol) was added, followed by benzyl bromide (6.44 ml, 54.1 mmol). The mixture was left overnight at rt, than concentrated under reduced pressure and partitioned between 1 N

15   HCl and ether. Organic phase was washed with saturated NaHCO$_3$, dried over Na$_2$SO$_4$ and evaporated . Flash chromatography using 20-30% gradient of ethyl acetate in hexanes afforded benzyl ester as a white powder (14.1 g, 93.4%). ^{1}H-NMR spectral data were in accordance with the published values.[33]

12'-Benzyl   hydroxydodecanoyl-2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-β-D-

20   **galactopyrano-se** (6). 1-Chloro sugar **4** (4.26 g, 11.67 mmol)and benzyl 12-hydroxydodecanoate (5) (4.3 g, 13.03 mmol) were dissolved in nitromethane-toluene 1:1 (122 ml) under argon and Hg(CN)$_2$ (3.51 g, 13.89 mmol) and powdered molecular sieves 4A (1.26 g) were added. The mixture was stirred at rt for 24 h, filtered and the filtrate concentrated under reduced pressure. The residue was partitioned between

25   dichloromethane and brine, organic layer was washed with brine, followed by 0.5 M KBr, dried (Na$_2$SO$_4$) and evaporated to a syrup. Flash silica gel column chromatography using 15-30% gradient of acetone in hexanes yielded product **6** as a colorless foam (6 g, 81%). ^{1}H-NMR δ 7.43 (m, 5H, phenyl), 5.60 (d, 1H, J$_{NH,2}$=8.8, NH), 5.44 (d, J$_{4,3}$=3.2, 1H, H4), 5.40 (dd, J$_{3,4}$=3.2, J$_{3,2}$=10.8, 1H, H3), 5.19 (s, 2H, *CH*$_2$Ph), 4.80 (d, J$_{1,2}$=8.0, 1H, H1), 4.23

30   (m, 2H, CH$_2$), 3.99 (m, 3H, H2, H6), 3.56 (m, 1H, H5), 2.43 (t, J=7.2, 2H,CH$_2$), 2.22 (s, 3H, Ac), 2.12 (s, 3H, Ac), 2.08 (s, 3H, Ac), 2.03 (s, 3H, Ac), 1.64 (m, 4H, 2×CH$_2$), 1.33 (br m, 14H, 7×CH$_2$). MS/ESI$^-$ m/z 634.5 (M-H)$^-$.

12'-Hydroxydodecanoyl-2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-β-D-galactopyranose (7).

35   Conjugate **6** (2 g, 3.14 mmol)) was dissolved in ethanol (50 ml) and 5% Pd-C (0.3 g) was added. The reaction mixture was hydrogenated overnight at 45 psi H$_2$, the catalyst was

filtered off and the filtrate evaporated to dryness to afford pure **7** (1.7 g, quantitative) as a white foam. ^{1}H-NMR $\delta\bullet$ 5.73 (d, 1H, $J_{NH,2}$=8.4, NH), 5.44 (d, $J_{4,3}$=3.0, 1H, H4), 5.40 (dd, $J_{3,4}$=3.0, $J_{3,2}$=11.2,1H, H3), 4.78 (d, $J_{1,2}$=8.8, 1H, H1), 4.21(m, 2H, CH_2), 4.02 (m, 3H, H2, H6), 3.55 (m, 1H, H5), 2.42 (m, 2H, CH_2), 2.23(s, 3H, Ac), 2.13 (s, 3H, Ac), 2.09 (s,
5   3H, Ac), 2.04 (s, 3H, Ac), 1.69 (m, 4H, 2xCH_2), 1.36 (br m, 14H, 7xCH_2). MS/ESI$^-$ m/z 544.0 (M-H)$^-$.

2'-(*N*-α,ε-bis-(12'-Hydroxydodecanoyl-2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-β-D-galac-topyranose)-L-lysyl)amino-2'-deoxy-5'-*O*-4,4'-dimethoxytrityl uridine (9). **7** (1.05 g, 1.92 mmol) was dissolved in anhydrous THF and *N*-hydroxysuccinimide (0.27 g,
10   2.35 mmol) and 1,3-dicyclohexylcarbodiimide (0.55 g, 2.67 mmol) were added. The reaction mixture was stirred at rt overnight, then filtered through Celite pad and the filtrate concentrated under reduced pressure. The crude NHSu ester **8** was dissolved in dry DMF (13 ml) containing diisopropylethylamine (0.67 ml, 3.85 mmol) and to this solution nucleoside **2** (0.64 g, 0.95 mmol was added). The reaction mixture was stirred at rt
15   overnight and than concentrated under reduced pressure. The residue was partitioned between water and dichloromethane, the aqueous layer extracted with dichloromethane, the organic layers combined, dried (Na_2SO_4) and evaporated to a syrup. Flash silica gel column chromatography using 2-3% gradient of methanol in ethyl acetate yielded **9** as a colorless foam (1.04 g, 63%). ^{1}H-NMR δ 7.42 (d, $J_{6,5}$=8.4, 1H, H6 Urd), 7.53-6.97 (m,
20   13H, aromatic), 6.12 (d, $J_{1',2'}$=8.0, 1H, H-1'), 5.41 (m, 3H, H5 Urd, H4 NAcGal), 5.15 (dd, $J_{3,4}$=3.6, $J_{3,2}$=11.2, 2H, H3 NAcGal), 4.87 (dd, $J_{2',3'}$=5.6, $J_{2',1'}$=8.0, 1H, H2'), 4.63 (d, $J_{1,2}$=8.0, 2H, H1 NAcGal), 4.42 (d, $J_{3',2'}$=5.6, 1H, H3'), 4.29-4.04 (m, 9H, H4', H2 NAcGal, H5 NacGal, CH_2), 3.95-3.82 (m, 8H, H6 NAcGal, 2xOMe), 3.62-3.42 (m, 4H, H5', H6 NAcGal), 3.26 (m, 2H, CH_2), 2.40-1.97 (m, 28H, CH_2, Ac), 1.95-1.30 (m, 50H,
25   CH_2). MS/ESI$^-$ m/z 1727.0 (M-H)$^-$.

2'-(*N*-α,ε-bis-(12'-Hydroxydodecanoyl-2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-β-D-galac-topyranose)-L-lysyl)amino-2'-deoxy-5'-*O*-4,4'-dimethoxytrityl uridine 3'-*O*-(2-cyanoethyl *N*,*N*-diisopropylphosphoramidite) (10). Conjugate **9** (0.87 g, 0.50 mmol) was dissolved in dry dichloromethane (10 ml) under argon and diisopropylethylamine
30   (0.36 ml, 2.07 mmol) and 1-methylimidazole (21 μL, 0.26 mmol) were added. The solution was cooled to 0 ^{o}C and 2-cyanoethyl diisopropylchlorophosphoramidite (0.19 ml, 0.85 mmol) was added. The reaction mixture was stirred at rt for 1 hour, than cooled to 0 ^{o}C and quenched with anhydrous ethanol (0.5 ml). After stirring for 10 min the solution was concentrated under reduced pressure (40 ^{o}C) and the residue dissolved in
35   dichloromethane and chromatographed on the column of silica gel using hexanes-ethyl acetate 1:1, followed by ethyl acetate and finally ethyl acetate-acetone 1:1 (1%

triethylamine was added to solvents) to afford the phosphoramidite **10** (680 mg, 69%). ^{31}P-NMR δ 152.0 (s), 149.3 (s). MS/ESI⁻ m/z 1928.0 (M-H)⁻.

N-(12'-Hydroxydodecanoyl-2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-β-D-galactopyranose)-D-threoninol (11). 12'-Hydroxydodecanoyl-2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-β-D-galac-topyranose **7** (850 mg, 1.56 mmol) was dissolved in DMF (5 ml) and to the solution N-hydroxysuccinimide (215 mg, 1.87 mmol) and 1,3-dicyclohexylcarbodimide (386 mg, 1.87 mmol) were added. The reaction mixture was stirred at rt overnight, the precipitate was filtered off and to the filtrate D-threoninol (197 mg, 1.87 mmol) was added. The mixture was stirred at rt overnight and concentrated in vacuo. The residue was partitioned between dichloromethane and 5% NaHCO₃, the organic layer was washed with brine, dried (Na₂SO₄) and evaporated to a syrup. Silica gel column chromatography using 1-10% gradient of methanol in dichloromethane afforded **11** as a colorless oil (0.7 g, 71%). ^{1}H-NMR δ 6.35 (d, J=7.6, 1H, NH), 5.77 (d, J=8.0, 1H, NH), 5.44 (d, $J_{4,3}$=3.6, 1H, H4), 5.37 (dd, $J_{3,4}$=3.6, $J_{3,2}$=11.2, 1H, H3), 4.77 (d, $J_{1,2}$=8.0, 1H, H1), 4.28-4.18 (m, 3H, CH₂, CH), 4.07-3.87 (m, 6H), 3.55 (m, 1H, H5), 3.09 (d, J=3.2, 1H, OH), 3.02 (t, J=4.6, 1H, OH), 2.34 (t, J=7.4 2H, CH₂), 2.23 (s, 3H, Ac), 2.10 (s, 3H, Ac), 2.04 (s, 3H, Ac), 1.76-1.61 (m, 2×CH₂), 1.35 (m, 14H, 7×CH₂), 1.29 (d, J=6.4, 3H, CH₃). MS/ESI⁻ m/z (M-H)⁻.

1-O-(4-Monomethoxytrityl)-N-(12'-hydroxydodecanoyl-2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-β-D-galactopyranose)-D-threoninol (12). To the solution of **11** (680 mg, 1.1 mmol) in dry pyridine (10 ml) p-anisylchlorotriphenylmethane (430 mg, 1.39 mmol) was added and the rection mixture was stirred, protected from moisture, overnight. Methanol (3 ml) was added and the solution stirred for 15 min and evaporated in vacuo. The residue was partitioned between dichloromethane and 5% NaHCO₃, the organic layer was washed with brine, dried (Na₂SO₄) and evaporated to a syrup. Silica gel column chromatography using 1-3% gradient of methanol in dichloromethane afforded **12** as a white foam (0.75 g, 77%). ^{1}H-NMR δ 7.48-6.92 (m, 14 H, aromatic), 6.15 (d, J=8.8, 1H, NH), 5.56 (d, J=8.0, 1H, NH), 5.45 (d, $J_{4,3}$=3.2, 1H, H4), 5.40 (dd, $J_{3,4}$=3.2, $J_{3,2}$=11.2, 1H, H3), 4.80 (d, $J_{1,2}$=8.0, 1H, H1), 4.3-4.13 (m, 3H, CH₂, CH), 4.25-3.92 (m, 4H, H6, H2, CH), 3.89 (s, 3H, OMe), 3.54 (m, 2H, H5, CH), 3.36 (dd, J=3.4, J=9.8, 1H, CH), 3.12 (d, J=2.8, 1H, OH), 2.31 (t, J=7.6, 2H, CH₂), 2.22 (s, 3H, Ac), 2.13 (s, 3H, Ac), 2.03 (s, 3H, Ac), 1.80-1.55 (m, 2×CH₂), 1.37 (m, 14H, 7×CH₂), 1.21 (d, J=6.4, 3H, CH₃). MS/ESI⁻ m/z 903.5 (M-H)⁻.

1-O-(4-Monomethoxytrityl)-N-(12'-hydroxydodecanoyl-2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-β-D-galactopyranose)-D-threoninol   3-O-(2-cyanoethyl   N,N-diisopropylphosphorami-dite) (13). Conjugate **12** (1.2 g, 1.33 mmol) was dissolved in

dry dichloromethane (15 ml) under argon and diisopropylethylamine (0.94 ml, 5.40 mmol) and 1-methylimidazole (55 µL, 0.69 mmol) were added. The solution was cooled to 0 $^\circ$C and 2-cyanoethyl N,N-diisopropyl-chlorophosphoramidite (0.51 ml, 2.29 mmol) was added. The reaction mixture was stirred at rt for 2 hours, than cooled to 0 $^\circ$C and

5 quenched with anhydrous ethanol (0.5 ml). After stirring for 10 min. the solution was concentrated under reduced pressure (40 $^\circ$C) and the residue dissolved in dichloromethane and chromatographed on the column of silica gel using 50-80% gradient of ethyl acetate in hexanes (1% triethylamine) to afford the phosphoramidite **13** (1.2 g, 82%). ^{31}P-NMR δ •149.41 (s), 149.23 (s).

10 Oligonucleotide synthesis

Phosphoramidites **10**, and **13**, were used along with standard 2'-O-TBDMS and 2'-O-methyl nucleoside phosphoramidites. Synthesis were conducted on a 394 (ABI) synthesizer using modified 2.5 µmol scale protocol with a 5 min coupling step for 2'-O-TBDMS protected nucleotides and 2.5 min coupling step for 2'-O-methyl nucleosides.

15 Coupling efficiency for the phosphoramidite **10** was lower than 50% while coupling efficiencies for phosphoramidite **13** was typically greater than 95% based on the measurement of released trityl cations. Once the synthesis was completed, the oligonucleotides were deprotected. The 5'-trityl groups were left attached to the oligomers to assist purification. Cleavage from the solid support and the removal of the

20 protecting groups was performed as described herein with the exception of using 20% piperidine in DMF for 15 min for the removal of Fm protection prior methylamine treatment. The 5'-tritylated oligomers were separated from shorter (trityl-off) failure sequences using a short column of SEP-PAK C-18 adsorbent. The bound, tritylated oligomers were detritylated on the column by treatment with 1% trifluoroacetic acid,

25 neutralized with triethylammonium acetate buffer, and than eluted. Further purification was achieved by reverse-phase HPLC. An example of a N-acetyl-D-galactosamine conjugate that can be synthesized using phosphoramidite **13** is shown in **Figure 15**.

Structures of the ribozyme conjugates were confirmed by MALDI-TOF MS.

Monomer synthesis

30 **2'-Amino-2'-deoxyuridine-*N*-acetyl-D-galactosamine conjugate**. The bis-Fmoc protected lysine linker was attached to the 2'-amino group of 2'-amino-2'-deoxyuridine using the EEDQ catalyzed peptide coupling. The 5'-OH was protected with 4,4'-dimethoxytrityl group to give **1**, followed by the cleavage of *N*-Fmoc groups with diethylamine to afford synthon **2** in the high overall yield.

2-acetamido-3,4,6-tetra-O-acetyl-1-chloro-D-galactopyranose **4** was synthesized with minor modifications according to the reported procedure (Findeis *supra*). Mercury salt catalyzed glycosylation of **4** with the benzyl ester of 12-hydroxydodecanoic acid **5** afforded glycoside **6** in 81 % yield. Hydrogenolysis of benzyl protecting group yielded **7** in a quantitative yield. The coupling of the sugar derivative with the nucleoside synthon was achieved through preactivation of the carboxylic function of **7** as *N*-hydroxysuccinimide ester **8**, followed by coupling to lysyl-2'-aminouridine conjugate **2**. The final conjugate **9** was than phosphitylated under standard conditions to afford the phosphoramidite **10** in 69% yield.

D-Threoninol-*N*-acetyl-D-galactosamine conjugate Using the similar strategy as described above, D-threoninol was coupled to **7** to afford conjugate **11** in a good yield. Monomethoxytritylation, followed by phosphitylation yielded the desired phosphoramidite **13**.

Example 2: Synthesis of Oxime linked nucleic acid/peptide conjugates (**Figures 16 and 17**)

12-Hydroxydodecanoic acid benzyl ester Benzyl bromide (10.28 ml, 86.45 mmol) was added dropwise to a solution of 12-hydroxydodecanoic acid (17 g, 78.59 mmol) and DBU (12.93 ml, 86.45 mmol) in absolute DMF (120 ml) under vigorous stirring at 0•C. After completeion of the addition reaction mixture was warmed to a room temperature and left overnight under stirring. TLC (hexane-ethylacetate 3:1) indicated complete transformation of the starting material. DMF was removed under reduced pressure and the residue was partitioned between ethyl ether and 1N HCl. Organic phase was separated, washed with saturated aq sodium bicarbonate and dried over sodium sulfate. Sodium sulfate was filtered off, filtrate was evaporated to dryness. The residue was crystallized from hexane to give 21.15 g (92%) of the title compound as a white powder.

12-O-N-Phthaloyl-dodecanoic acid benzyl ester (15). Diethylazodicarboxylate (DEAD, 16.96 ml, 107.7 mmol) was added dropwise to the mixture of 12-Hydroxydodecanoic acid benzyl ester (21g, 71.8 mmol), triphenylphosphine (28.29 g, 107.7 mmol) and N-hydroxyphthalimide (12.88 g, 78.98 mmol) in absolute THF (250 ml) at –20• – -30•C under stirring. The reaction mixture was stirred at this temperature for additional 2-3h, after which time TLC (hexane-ethylacetate 3:1) indicated reaction completion. The solvent was removed in *vacuo* and the residue was treated ether (250 ml). Formed precipitate of triphenylphosphine oxide was filtered off, mother liquor was evaporated to dryness and the residue was dissolved in methylene chloride and purified by flash chromatography on silica gel in hexane-ethyl acetate (7:3). Appropriate fractions were pooled and evaporated to dryness to afford 26.5 g(84.4%) of compound **15**.

12-O-N-Phthaloyl-dodecanoic acid (16). Compound **15** (26.2 g, 59.9 mmol) was dissolved in 225 ml of ethanol-ethylacetate (3.5:1) mixture and 10% Pd/C (2.6 g) was added. The reaction mixture was hydrogenated in Parr apparatus for 3 hours. Reaction mixture was filtered through celite and evaporated to dryness. The residue was crystallized from methanol to provide 15.64 g (75%) of compound **16**.

12-O-N-Phthaloyl-dodecanoic acid 2,3-di-hydroxy-propylamide (18) The mixture of compound **16** (15.03 g, 44.04 mmol), dicyclohexylcarbodiimide (10.9 g, 52.85 mmol) and N-hydroxysuccinimide (6.08 g, 52.85 mmol) in absolute DMF (150 ml) was stirred at room temperature overnight. TLC (methylene chloride –methanol 9:1) indicated complete conversion of the starting material and formation of NHS ester **17**. Then aminopropanediol (4.01 g, 44 mmol) was added and the reaction mixture was stirred at room temperature for another 2 h. The formed precipitate of dicyclohexylurea was removed by filtration, filtrate was evaporated under reduced pressure. The residue was partitioned between ethyl acetate and saturated aq sodium bicarbonate. The whole mixture was filtered to remove any insoluble material and clear layers were separated. Organic phase was concentrated in vacuo until formation of crystalline material. The precipitate was filtered off and washed with cold ethylacetate to produce 10.86 g of compound **17**. Combined mother liquor and washings were evaporated to dryness and crystallized from ethylacetate to afford 3.21 g of compound **18**. Combined yield – 14.07 g (73.5%).

12-O-N-Phthaloyl-dodecanoic acid 2-hydroxy,3-dimethoxytrityloxy-propylamide (19) Dimethoxytrityl chloride (12.07 g, 35.62 mmol) was added to a stirred solution of compound **18** (14.07 g, 32.38 mmol) in absolute pyridine (130 ml) at 0•C. The reaction solution was kept at 0•C overnight. Then it was quenched with MeOH (10 ml) and evaporated to dryness. The residue was dissolved in methylene chloride and washed with saturated aq sodium bicarbonate. Organic phase was separated, dried over sodium sulfate and evaporated to dryness. The residue was purified by flash chromatography on silica gel using step gradient of acetone in hexanes (3:7 to 1:1) as an eluent. Appropriate fractions were pooled and evaporated to provide 14.73g (62%) of compound **19**, as a colorless oil.

12-O-N-Phthaloyl-dodecanoic acid 2-O-(cyanoethyl-N.N-diisopropylamino-

phosphoramidite),3-dimethoxytrityloxy-propylamide (20). Phosphitylated according to Sanghvi, *et al.,* 2000, *Organic Process Research and Development*, 4, 175-81.

Purified by flash chromatography on silica gel using step gradient of acetone in hexanes (1:4 to 3:7) containing 0.5% of triethylamine. Yield – 82%, colourless oil.

<u>Oxidation of peptides</u>

Peptide (3.3 mg, 3.3 μmol) was dissolved in 10mM AcONa and 2 eq of sodium periodate (100mM soln in water) was added. Final reaction volume – 0.5 ml. After 10 minutes reaction mixture was purified using analytical HPLC on Phenomenex Jupiter 5u C18 300A (150x4.6 mm) column; solvent A: 50mM KH$_2$PO$_4$ (pH 3); solvent B: 30% of solvent A in MeCN; gradient B over 30 min. Appropriate fractions were pooled and concentrated on a SpeedVac to dryness. Yield: quantitative.

<u>Conjugation reaction of Herzyme-ONH2-linker with N-glyoxyl peptide (**Figure 17**)</u>

Herzyme (SEQ ID NO: 13) with a 5'-terminal linker (100 OD) was mixed with oxidized peptide (3-5 eq) in 50 mM KH2PO4 (pH3, reaction volume 1 ml) and kept at room temperature for 24-48h. The reaction mixture was purified using analytical HPLC on a Phenomenex Jupiter 5u C18 300A (150x4.6 mm) column; solvent A: 10mM TEAA; solvent B: 10mM TEAA/MeCN. Appropriate fractions were pooled and concentrated on a SpeedVac to dryness to provide desired conjugate. ESMS: calculated: 12699, determined: 12698.

<u>Example 5: Synthesis of Phospholipid enzymatic nucleic acid conjugates (**Figure 19**)</u>

A phospholipid enzymatic nucleic acid conjugate (see **Figure 19**) was prepared by coupling a C18H37 phosphoramidite to the 5'-end of an enzymatic nucleic acid molecule (Angiozyme™, SEQ ID NO: 24) during solid phase oligonucleotide synthesis on an ABI 394 synthesizer using standard synthesis chemistry. A 5'-terminal linker comprising 3'-AdT-di-Glycerol-5', where A is Adenosine, dT is 2'-deoxy Thymidine, and di-Glycerol is a di-DMT-Glycerol linker (Chemgenes CAT number CLP-5215), is used to attach two C18H37 phosphoramidites to the enzymatic nucleic acid molecule using standard synthesis chemistry. Additional equivalents of the C18H37 phosphoramidite were used for the bis-coupling. Similarly, other nucleic acid conjugates as shown in **Figure 18** can be prepared according to similar methodology.

<u>Example 6: Synthesis of PEG enzymatic nucleic acid conjugates (**Figure 20**)</u>

A 40K-PEG enzymatic nucleic acid conjugate (see **Figure 20**) was prepared by post synthetic N-hydroxysuccinimide ester coupling of a PEG derivative (Shearwater Polymers Inc, CAT number PEG2-NHS) to the 5'-end of an enzymatic nucleic acid molecule (Angiozyme™, SEQ ID NO: 24). A 5'-terminal linker comprising 3'-AdT-C6-amine-5', where A is Adenosine, dT-C6-amine is 2'-deoxy Thymidine with a C5 linked six carbon amine linker (Glen Research CAT number 10-1039-05), is used to attach the PEG derivative to the enzymatic nucleic acid molecule using NHS coupling chemistry.

Angiozyme™ with the C6dT-NH2 at the 5' end was synthesized and deprotected using standard oligonucleotide synthesis procedures as described herein. The crude

sample was subsequently loaded onto a reverse phase column and rinsed with sodium chloride solution (0.5 M). The sample was then desalted with water on the column until the concentration of sodium chloride was close to zero. Acetonitrile was used to elute the sample from the column. The crude product was then concentrated and lyophilized to

5   dryness.

The crude material (Angiozyme™) with 5'-amino linker (50 mg) was dissolved in sodium borate buffer (1.0 mL, pH 9.0). The PEG NHS ester (200 mg) was dissolved in anhydrous DMF (1.0 mL). The Angiozyme™ buffer solution was then added to the PEG NHS ester solution. The mixture was immediately vortexed for 5 minutes. Sodium acetate

10   buffer solution (5 mL, pH 5.2) was used to quench the reaction. Conjugated material was then purified by ion-exchange and reverse phase chromatography.

Example 7: Phamacokinetics of PEG ribozyme acid conjugate (**Figure 21**)

Forty-eight female C57Bl/6 mice were given a single subcutaneous (SC) bolus of 30 mg/kg Angiozyme™ and 30 mg/kg Angiozyme™/40K PEG conjugate. Plasma was

15   collected out to 24 hours post ribozyme injection. Plasma samples were analyzed for full length ribozyme by a hybridization assay.

Oligonucleotides complimentary to the 5' and 3' ends of Angiozyme™ were synthesized with biotin at one oligo, and FITC on the other oligo. A biotin oligo and FITC labeled oligo pair are incubated at 1 ug/ml with known concentrations of

20   Angiozyme™ at 75degrees C for 5 min. After 10 minutes at RT, the mixture is allowed to bind to streptavidin coated wells of a 96-wll plate for two hours. The plate is washed with Tris-saline and detergent, and peroxidase labeled anti-FITC antibody is added. After one hour, the wells are washed, and the enzymatic reaction is developed, then read on an ELISA plate reader. Results are shown in **Figure 21**.

25

Example 8: Phamacokinetics of Phospholipid ribozyme conjugate (**Figure 22**)

Seventy-two female C57Bl/6 mice were given a single intravenous (4) bolus of 30 mg/kg Angiozyme™ and 30 mg/kg Angiozyme™ conjugated with phospholipid (**Figure 19**). Plasma was collected out to 3 hours post ribozyme injection. Plasma samples were

30   analyzed for full length ribozyme by a hybridization assay.

Oligonucleotides complimentary to the 5' and 3' ends of Angiozyme™ were synthesized with biotin at one oligo, and FITC on the other oligo. A biotin oligo and FITC labeled oligo pair are incubated at 1 ug/ml with known concentrations of Angiozyme™ at 75degrees C for 5 min. After 10 minutes at RT, the mixture is allowed

35   to bind to streptavidin coated wells of a 96-wll plate for two hours. The plate is washed with Tris-saline and detergent, and peroxidase labeled anti-FITC antibody is added. After

one hr, the wells are washed, and the enzymatic reaction is developed, then read on an ELISA plate reader. Results are shown in **Figure 22**.

Example 9:   Synthesis of Protein or Peptide conjugates with biodegradable linkers
(Figures 24-26, and 29)

Proteins and peptides can be conjugated with various molecules, including PEG, via biodegradable nucleic acid linker molecules of the invention, using oxime and morpholino linkages. For example, a therapeutic antibody can be conjugated with PEG to improve the **Figure 24** shows a non-limiting example of a synthetic approach for synthesizing peptide or protein conjugates to PEG utilizing a biodegradable linker, the example shown is for a protein conjugate. Other conjugates can be synthesized in a similar manner where the protein or peptide is conjugated to molecules other than PEG, such as small molecules, toxins, radioisotopes, peptides or other proteins. (a) The protein of interest, such as an antibody or interferon, is synthesized with a terminal Serine or Threonine moiety that is oxidized, for example with sodium periodate. The oxidized protein is then coupled to a nucleic acid linker molecule that is designed to be biodegradable, for example a cytidine-deoxythymidine, cytidine-deoxyuridine, adenosine-deoxythymidine, or adenosine-deoxyuridine dimer that contains an oxyamino (O-NH$_2$) function. Other biodegradable nucleic acid linkers can be similarly used, for example other dimers, trimers, tetramers etc. that are designed to be biodegradable. The example shown makes use of a 5'-oxyamino moiety, however, other examples can utilize an oxyamino at other positions within the nucleic acid molecule, for example at the 2'-position, 3'-position, or at a nucleic acid base position. (b) The protein/nucleic acid conjugate is then oxidized to generate a dialdehyde function that is coupled to PEG molecule comprising an amino group (H$_2$N-PEG), for example a PEG molecule with an amino linker.   Other amino containing molecules can be conjugated as shown in the figure, for example small molecules, toxins, or radioisotope labeled molecules.

Proteins and peptides can be conjugated with various molecules, including PEG, via biodegradable nucleic acid linker molecules of the invention, using oxime and phosphoramidate linkages. **Figure 25** shows a non-limiting example of a synthetic approach for synthesizing peptide or protein conjugates to PEG utilizing a biodegradable linker, the example shown is for a protein conjugate. Other conjugates can be synthesized in a similar manner where the protein or peptide is conjugated to molecules other than PEG, such as small molecules, toxins, radioisotopes, peptides or other proteins.   The protein of interest, such as an antibody or interferon, is synthesized with a terminal Serine or Threonine moiety that is oxidized, for example with sodium periodate. The oxidized

protein is then coupled to a nucleic acid linker molecule that is designed to be biodegradable, for example a cytidine-deoxythymidine, cytidine-deoxyuridine, adenosine-deoxythymidine, or adenosine-deoxyuridine dimer that contains an oxyamino (O-NH_2) function and a terminal phosphate group. Terminal phosphate groups can be introduced

5   during synthesis of the nucleic acid molecule using chemical phosphorylation reagents, such as Glen Research Cat Nos. 10-1909-02, 10-1913-02, 10-1914-02, and 10-1918-02. Other biodegradable nucleic acid linkers can be similarly used, for example other dimers, trimers, tetramers etc. that are designed to be biodegradable. The example shown makes use of a 5'-oxyamino moiety, however, other examples can utilize an oxyamino at other

10   positions within the nucleic acid molecule, for example at the 2'-position, 3'-position, or at a nucleic acid base position. The protein/nucleic acid conjugate terminal phosphate group is then activated with an activator reagent, such as NMI and/or tetrazole, and coupled a PEG molecule comprising an amino group (H_2N-PEG), for example a PEG molecule with an amino linker. Other amino containing molecules can be conjugated as

15   shown in the figure, for example small molecules, toxins, or radioisotope labeled molecules.

Proteins and peptides can be conjugated with various molecules, including PEG, via biodegradable nucleic acid linker molecules of the invention, using phosphoramidate linkages. **Figure 26** shows a non-limiting example of a synthetic approach for

20   synthesizing peptide or protein conjugates to PEG utilizing a biodegradable linker, the example shown is for a protein conjugate. Other conjugates can be synthesized in a similar manner where the protein or peptide is conjugated to molecules other than PEG, such as small molecules, toxins, radioisotopes, peptides or other proteins. (a) A nucleic acid linker molecule that is designed to be biodegradable, for example a cytidine-

25   deoxythymidine, cytidine-deoxyuridine, adenosine-deoxythymidine, or adenosine-deoxyuridine dimer, is synthesized with a terminal phosphate group. Other biodegradable nucleic acid linkers can be similarly used, for example other dimers, trimers, tetramers etc. that are designed to be biodegradable. The protein/nucleic acid conjugate terminal phosphate group is then activated with an activator reagent, such as NMI and/or tetrazole,

30   and coupled a PEG molecule comprising an amino group (H_2N-PEG), for example a PEG molecule with an amino linker. Other amino containing molecules can be conjugated as shown in the figure, for example small molecules, toxins, or radioisotope labeled molecules. The terminal protecting group, for example a dimethoxytrityl group, is removed from the conjugate and a terminal phosphite group is introduced with a

35   phosphitylating reagent, such as N,N-diisopropyl-2-cyanoethyl chlorophosphoramidite. (b) The PEG/nucleic acid conjugate is then coupled to a peptide or protein comprising an amino group, such as the amino terminus or amino side chain of a suitably protected

peptide or protein or via an amino linker. The conjugate is then oxidized and any protecting groups are removed to yield the protein/PEG conjugate comprising a biodegradable linker.

5   Proteins and peptides can be conjugated with various molecules, including PEG, via biodegradable nucleic acid linker molecules of the invention, using phosphoramidate linkages from coupling protein-based phosphoramidites. **Figure 29** shows a non-limiting example of a synthetic approach for synthesizing peptide or protein conjugates to PEG utilizing a biodegradable linker, the example shown is for a protein conjugate. Other conjugates can be synthesized in a similar manner where the protein or peptide is
10   conjugated to molecules other than PEG, such as small molecules, toxins, radioisotopes, peptides or other proteins. The protein of interest, such as an antibody or interferon, is synthesized with a terminal Serine, Threonin, or Tyrosine moiety that is phosphitylated, for example with N,N-diisopropyl-2-cyanoethyl chlorophosphoramidite. The phosphitylated protein is then coupled to a nucleic acid linker molecule that is designed to
15   be biodegradable, for example a cytidine-deoxythymidine, cytidine-deoxyuridine, adenosine-deoxythymidine, or adenosine-deoxyuridine dimer that contains conjugated PEG molecule as described in Figure 18. Other biodegradable nucleic acid linkers can be similarly used, for example other dimers, trimers, tetramers etc. that are designed to be biodegradable.

20   Example 10: Galactosamine ribozyme conjugate targeting HBV

A nuclease-resistance ribozyme directed against the Heptatitis B viral RNA (HBV) (HepBzyme™) is in early stages of preclinical development. HepBzyme, which targets site 273 of the Hepatitis B viral RNA, has produced statistically significant decreases in serum HBV levels in a HBV transgenic mouse model in a dose-dependent manner (30
25   and 100 mg/kg/day). In an effort to improve hepatic uptake by targeting the asialoglycoprotein receptor, a series of 5 branched galactosamine residues were attached via phosphate linkages to the 5'-terminus of HepBzyme (Gal-HepBzyme). The affect of the galactosamine conjugation on HepBzyme was assessed by quantitation of ^{32}P-labeled HepBzyme and Gal-HepBzyme in plasma, liver and kidney of mice following a single SC
30   bolus administration of 30 mg/kg. The plasma disposition of the intact ribozyme was similar between Gal-HepBzyme and HepBzyme. An approximate three-fold increase in the maximum observed concentration of intact ribozyme in liver (C_{max}) was observed in liver for Gal-HepBzyme (6.1 ± 1.8 ng/mg) vs. HepBzyme (2.2 ± 0.8 ng/mg) ($p < 0.05$). The area under the curve (AUCall) for Gal-HepBzyme was also increased by
35   approximately two-fold. This was accompanied by a substantial decrease (approximately

40%) in the AUC_{all} for intact ribozyme in kidney. In addition to the significant increase in C_{max} observed for intact Gal-HepBzyme in the liver, there was an increase in the total number of ribozyme equivalents, which may be suggestive of increased affinity of both the intact ribozyme and metabolites for asialoglycoprotein receptor and galactose-specific receptors in the liver. These data demonstrate that conjugation of a ribozyme with galactosamine produces a compound with a more favorable disposition profile, and illustrates the utility of conjugated ribozymes with improved *in vivo* pharmacokinetics and biodistribution.

One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The methods and compositions described herein are exemplary and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art, which are encompassed within the spirit of the invention, are defined by the scope of the claims.

It will be readily apparent to one skilled in the art that varying substitutions and modifications can be made to the invention disclosed herein without departing from the scope and spirit of the invention. Thus, such additional embodiments are within the scope of the present invention and the following claims.

The invention illustratively described herein suitably can be practiced in the absence of any element or elements, limitation or limitations which is not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising", "consisting essentially of" and "consisting of" may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is no intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by various embodiments, optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention as defined by the description and the appended claims.

In addition, where features or aspects of the invention are described in terms of Markush groups or other grouping of alternatives, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group or other group.

Other embodiments are within the following claims.

TABLE I

Characteristics of naturally occurring ribozymes

Group I Introns

- Size: ~150 to >1000 nucleotides.

- Requires a U in the target sequence immediately 5' of the cleavage site.

- Binds 4-6 nucleotides at the 5'-side of the cleavage site.

- Reaction mechanism: attack by the 3'-OH of guanosine to generate cleavage products with 3'-OH and 5'-guanosine.

- Additional protein cofactors required in some cases to help folding and maintenance of the active structure.

- Over 300 known members of this class. Found as an intervening sequence in *Tetrahymena thermophila* rRNA, fungal mitochondria, chloroplasts, phage T4, blue-green algae, and others.

- Major structural features largely established through phylogenetic comparisons, mutagenesis, and biochemical studies [i,ii].

- Complete kinetic framework established for one ribozyme [iii,iv,v,vi].

- Studies of ribozyme folding and substrate docking underway [vii,viii,ix].

- Chemical modification investigation of important residues well established [x,xi].

- The small (4-6 nt) binding site may make this ribozyme too non-specific for targeted RNA cleavage, however, the Tetrahymena group I intron has been used to repair a "defective" β-galactosidase message by the ligation of new β-galactosidase sequences onto the defective message [xii].

RNAse P RNA (M1 RNA)
- Size: ~290 to 400 nucleotides.

- RNA portion of a ubiquitous ribonucleoprotein enzyme.

- Cleaves tRNA precursors to form mature tRNA [xiii].

- Reaction mechanism: possible attack by M^{2+}-OH to generate cleavage products with 3'-OH and 5'-phosphate.

- RNAse P is found throughout the prokaryotes and eukaryotes. The RNA subunit has been sequenced from bacteria, yeast, rodents, and primates.

- Recruitment of endogenous RNAse P for therapeutic applications is possible through hybridization of an External Guide Sequence (EGS) to the target RNA [xiv, xv]

- Important phosphate and 2' OH contacts recently identified [xvi, xvii]

Group II Introns
- Size: >1000 nucleotides.

- Trans cleavage of target RNAs recently demonstrated [xviii, xix].

- Sequence requirements not fully determined.

- Reaction mechanism: 2'-OH of an internal adenosine generates cleavage products with 3'-OH and a "lariat" RNA containing a 3'-5' and a 2'-5' branch point.

- Only natural ribozyme with demonstrated participation in DNA cleavage [xx, xxi] in addition to RNA cleavage and ligation.

- Major structural features largely established through phylogenetic comparisons [xxii].

- Important 2' OH contacts beginning to be identified [xxiii]

- Kinetic framework under development [xxiv]

Neurospora VS RNA
- Size: ~144 nucleotides.

- Trans cleavage of hairpin target RNAs recently demonstrated [xxv].

- Sequence requirements not fully determined.

- Reaction mechanism: attack by 2'-OH 5' to the scissile bond to generate cleavage products with 2',3'-cyclic phosphate and 5'-OH ends.

- Binding sites and structural requirements not fully determined.

- Only 1 known member of this class. Found in Neurospora VS RNA.

Hammerhead Ribozyme
(see text for references)

- Size: ~13 to 40 nucleotides.

- Requires the target sequence UH immediately 5' of the cleavage site.

- Binds a variable number nucleotides on both sides of the cleavage site.

- Reaction mechanism: attack by 2'-OH 5' to the scissile bond to generate cleavage products with 2',3'-cyclic phosphate and 5'-OH ends.

- 14 known members of this class. Found in a number of plant pathogens (virusoids) that use RNA as the infectious agent.

- Essential structural features largely defined, including 2 crystal structures [xxvi, xxvii]

- Minimal ligation activity demonstrated (for engineering through *in vitro* selection) [xxviii]

- Complete kinetic framework established for two or more ribozymes [xxix].

- Chemical modification investigation of important residues well established [xxx].

Hairpin Ribozyme
- Size: ~50 nucleotides.

- Requires the target sequence GUC immediately 3' of the cleavage site.

- Binds 4-6 nucleotides at the 5'-side of the cleavage site and a variable number to the 3'-side of the cleavage site.

- Reaction mechanism: attack by 2'-OH 5' to the scissile bond to generate cleavage products with 2',3'-cyclic phosphate and 5'-OH ends.

- 3 known members of this class. Found in three plant pathogen (satellite RNAs of the tobacco ringspot virus, arabis mosaic virus and chicory yellow mottle virus) which uses RNA as the infectious agent.

- Essential structural features largely defined [xxxi, xxxii, xxxiii, xxxiv]

- Ligation activity (in addition to cleavage activity) makes ribozyme amenable to engineering through *in vitro* selection [xxxv]

- Complete kinetic framework established for one ribozyme [xxxvi].

- Chemical modification investigation of important residues begun [xxxvii, xxxviii].

Hepatitis Delta Virus (HDV) Ribozyme

- Size: ~60 nucleotides.

- Trans cleavage of target RNAs demonstrated [xxxix].

- Binding sites and structural requirements not fully determined, although no sequences 5' of cleavage site are required.   Folded ribozyme contains a pseudoknot structure [xl].

- Reaction mechanism: attack by 2'-OH 5' to the scissile bond to generate cleavage products with 2',3'-cyclic phosphate and 5'-OH ends.

- Only 2 known members of this class. Found in human HDV.

- Circular form of HDV is active and shows increased nuclease stability [xli]

---

i .   Michel, Francois; Westhof, Eric. Slippery substrates. Nat. Struct. Biol. (1994), 1(1), 5-7.

ii .   Lisacek, Frederique; Diaz, Yolande; Michel, Francois. Automatic identification of group I intron cores in genomic DNA sequences. J. Mol. Biol. (1994), 235(4), 1206-17.

iii .   Herschlag, Daniel; Cech, Thomas R.. Catalysis of RNA cleavage by the Tetrahymena thermophila ribozyme. 1. Kinetic description of the reaction of an RNA substrate complementary to the active site. Biochemistry (1990), 29(44), 10159-71.

iv .   Herschlag, Daniel; Cech, Thomas R.. Catalysis of RNA cleavage by the Tetrahymena thermophila ribozyme. 2. Kinetic description of the reaction of an RNA substrate that forms a mismatch at the active site. Biochemistry (1990), 29(44), 10172-80.

v .   Knitt, Deborah S.; Herschlag, Daniel. pH Dependencies of the Tetrahymena Ribozyme Reveal an Unconventional Origin of an Apparent pKa. Biochemistry (1996), 35(5), 1560-70.

vi .   Bevilacqua, Philip C.; Sugimoto, Naoki; Turner, Douglas H.. A mechanistic framework for the second step of splicing catalyzed by the Tetrahymena ribozyme. Biochemistry (1996), 35(2), 648-58.

vii .   Li, Yi; Bevilacqua, Philip C.; Mathews, David; Turner, Douglas H.. Thermodynamic and activation parameters for binding of a pyrene-labeled substrate by the Tetrahymena ribozyme: docking is not diffusion-controlled and is driven by a favorable entropy change. Biochemistry (1995), 34(44), 14394-9.

viii .   Banerjee, Aloke Raj; Turner, Douglas H.. The time dependence of chemical modification reveals slow steps in the folding of a group I ribozyme. Biochemistry (1995), 34(19), 6504-12.

ix .   Zarrinkar, Patrick P.; Williamson, James R.. The P9.1-P9.2 peripheral extension helps guide folding of the Tetrahymena ribozyme. Nucleic Acids Res. (1996), 24(5), 854-8.

x .   Strobel, Scott A.; Cech, Thomas R.. Minor groove recognition of the conserved G.cntdot.U pair at the Tetrahymena ribozyme reaction site. Science (Washington, D. C.) (1995), 267(5198), 675-9.

xi .   Strobel, Scott A.; Cech, Thomas R.. Exocyclic Amine of the Conserved G.cntdot.U Pair at the Cleavage Site of the Tetrahymena Ribozyme Contributes to 5'-Splice Site Selection and Transition State Stabilization. Biochemistry (1996), 35(4), 1201-11.

xii .   Sullenger, Bruce A.; Cech, Thomas R.. Ribozyme-mediated repair of defective mRNA by targeted trans-splicing. Nature (London) (1994), 371(6498), 619-22.

[xiii].   Robertson, H.D.; Altman, S.; Smith, J.D. J. Biol. Chem., 247, 5243-5251 (1972).

[xiv].   Forster, Anthony C.; Altman, Sidney. External guide sequences for an RNA enzyme. Science (Washington, D. C., 1883-) (1990), 249(4970), 783-6.

[xv].   Yuan, Y.; Hwang, E. S.; Altman, S. Targeted cleavage of mRNA by human RNase P. Proc. Natl. Acad. Sci. USA (1992) 89, 8006-10.

[xvi].   Harris, Michael E.; Pace, Norman R.. Identification of phosphates involved in catalysis by the ribozyme RNase P RNA. RNA (1995), 1(2), 210-18.

[xvii].   Pan, Tao; Loria, Andrew; Zhong, Kun. Probing of tertiary interactions in RNA: 2'-hydroxyl-base contacts between the RNase P RNA and pre-tRNA. Proc. Natl. Acad. Sci. U. S. A. (1995), 92(26), 12510-14.

[xviii].   Pyle, Anna Marie; Green, Justin B.. Building a Kinetic Framework for Group II Intron Ribozyme Activity: Quantitation of Interdomain Binding and Reaction Rate. Biochemistry (1994), 33(9), 2716-25.

[xix].   Michels, William J. Jr.; Pyle, Anna Marie. Conversion of a Group II Intron into a New Multiple-Turnover Ribozyme that Selectively Cleaves Oligonucleotides: Elucidation of Reaction Mechanism and Structure/Function Relationships. Biochemistry (1995), 34(9), 2965-77.

[xx].   Zimmerly, Steven; Guo, Huatao; Eskes, Robert; Yang, Jian; Perlman, Philip S.; Lambowitz, Alan M.. A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. Cell (Cambridge, Mass.) (1995), 83(4), 529-38.

[xxi].   Griffin, Edmund A., Jr.; Qin, Zhifeng; Michels, Williams J., Jr.; Pyle, Anna Marie. Group II intron ribozymes that cleave DNA and RNA linkages with similar efficiency, and lack contacts with substrate 2'-hydroxyl groups. Chem. Biol. (1995), 2(11), 761-70.

[xxii].   Michel, Francois; Ferat, Jean Luc. Structure and activities of group II introns. Annu. Rev. Biochem. (1995), 64, 435-61.

[xxiii].   Abramovitz, Dana L.; Friedman, Richard A.; Pyle, Anna Marie. Catalytic role of 2'-hydroxyl groups within a group II intron active site. Science (Washington, D. C.) (1996), 271(5254), 1410-13.

[xxiv].   Daniels, Danette L.; Michels, William J., Jr.; Pyle, Anna Marie. Two competing pathways for self-splicing by group II introns: a quantitative analysis of in vitro reaction rates and products. J. Mol. Biol. (1996), 256(1), 31-49.

[xxv].   Guo, Hans C. T.; Collins, Richard A.. Efficient trans-cleavage of a stem-loop RNA substrate by a ribozyme derived from Neurospora VS RNA. EMBO J. (1995), 14(2), 368-76.

[xxvi].   Scott, W.G., Finch, J.T., Aaron,K. The crystal structure of an all RNA hammerhead ribozyme:Aproposed mechanism for RNA catalytic cleavage. Cell, (1995), 81, 991-1002.

[xxvii].   McKay, Structure and function of the hammerhead ribozyme: an unfinished story. RNA, (1996), 2, 395-403.

[xxviii].   Long, D., Uhlenbeck, O., Hertel, K. Ligation with hammerhead ribozymes. US Patent No. 5,633,133.

[xxix].   Hertel, K.J., Herschlag, D., Uhlenbeck, O. A kinetic and thermodynamic framework for the hammerhead ribozyme reaction. Biochemistry, (1994) 33, 3374-3385.Beigelman, L., et al., Chemical modifications of hammerhead ribozymes. J. Biol. Chem., (1995) 270, 25702-25708.

[xxx].   Beigelman, L., et al., Chemical modifications of hammerhead ribozymes. J. Biol. Chem., (1995) 270, 25702-25708.

[xxxi].   Hampel, Arnold; Tritz, Richard; Hicks, Margaret; Cruz, Phillip. 'Hairpin' catalytic RNA model: evidence for helixes and sequence requirement for substrate RNA. Nucleic Acids Res. (1990), 18(2), 299-304.

[xxxii].   Chowrira, Bharat M.; Berzal-Herranz, Alfredo; Burke, John M.. Novel guanosine requirement for catalysis by the hairpin ribozyme. Nature (London) (1991), 354(6351), 320-2.

[xxxiii].   Berzal-Herranz, Alfredo; Joseph, Simpson; Chowrira, Bharat M.; Butcher, Samuel E.; Burke, John M.. Essential nucleotide sequences and secondary structure elements of the hairpin ribozyme. EMBO J. (1993), 12(6), 2567-73.

[xxxiv].   Joseph, Simpson; Berzal-Herranz, Alfredo; Chowrira, Bharat M.; Butcher, Samuel E.. Substrate selection rules for the hairpin ribozyme determined by in vitro selection, mutation, and analysis of mismatched substrates. Genes Dev. (1993), 7(1), 130-8.

[xxxv].   Berzal-Herranz, Alfredo; Joseph, Simpson; Burke, John M.. In vitro selection of active

hairpin ribozymes by sequential RNA-catalyzed cleavage and ligation reactions. Genes Dev. (1992), 6(1), 129-34.

[xxxvi]. Hegg, Lisa A.; Fedor, Martha J.. Kinetics and Thermodynamics of Intermolecular Catalysis by Hairpin Ribozymes. Biochemistry (1995), 34(48), 15813-28.

[xxxvii]. Grasby, Jane A.; Mersmann, Karin; Singh, Mohinder; Gait, Michael J.. Purine Functional Groups in Essential Residues of the Hairpin Ribozyme Required for Catalytic Cleavage of RNA. Biochemistry (1995), 34(12), 4068-76.

[xxxviii]. Schmidt, Sabine; Beigelman, Leonid; Karpeisky, Alexander; Usman, Nassim; Sorensen, Ulrik S.; Gait, Michael J.. Base and sugar requirements for RNA cleavage of essential nucleoside residues in internal loop B of the hairpin ribozyme: implications for secondary structure. Nucleic Acids Res. (1996), 24(4), 573-81.

[xxxix]. Perrotta, Anne T.; Been, Michael D.. Cleavage of oligoribonucleotides by a ribozyme derived from the hepatitis .delta. virus RNA sequence. Biochemistry (1992), 31(1), 16-21.

[xl]. Perrotta, Anne T.; Been, Michael D.. A pseudoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA. Nature (London) (1991), 350(6317), 434-6.

[xli]. Puttaraju, M.; Perrotta, Anne T.; Been, Michael D.. A circular trans-acting hepatitis delta virus ribozyme. Nucleic Acids Res. (1993), 21(18), 4253-8.

Table II:

A. 2.5 µmol Synthesis Cycle ABI 394 Instrument

Reagent	Equivalents	Amount	Wait Time* DNA	Wait Time* 2'-O-methyl	Wait Time*RNA
Phosphoramidites	6.5	163 µL	45 sec	2.5 min	7.5 min
S-Ethyl Tetrazole	23.8	238 µL	45 sec	2.5 min	7.5 min
Acetic Anhydride	100	233 µL	5 sec	5 sec	5 sec
N-Methyl Imidazole	186	233 µL	5 sec	5 sec	5 sec
TCA	176	2.3 mL	21 sec	21 sec	21 sec
Iodine	11.2	1.7 mL	45 sec	45 sec	45 sec
Beaucage	12.9	645 µL	100 sec	300 sec	300 sec
Acetonitrile	NA	6.67 mL	NA	NA	NA

B. 0.2 µmol Synthesis Cycle ABI 394 Instrument

Reagent	Equivalents	Amount	Wait Time* DNA	Wait Time* 2'-O-methyl	Wait Time*RNA
Phosphoramidites	15	31 µL	45 sec	233 sec	465 sec
S-Ethyl Tetrazole	38.7	31 µL	45 sec	233 min	465 sec
Acetic Anhydride	655	124 µL	5 sec	5 sec	5 sec
N-Methyl Imidazole	1245	124 µL	5 sec	5 sec	5 sec
TCA	700	732 µL	10 sec	10 sec	10 sec
Iodine	20.6	244 µL	15 sec	15 sec	15 sec
Beaucage	7.7	232 µL	100 sec	300 sec	300 sec
Acetonitrile	NA	2.64 mL	NA	NA	NA

C. 0.2 µmol Synthesis Cycle 96 well Instrument

Reagent	Equivalents:DNA/ 2'-O-methyl/Ribo	Amount: DNA/2'-O-methyl/Ribo	Wait Time* DNA	Wait Time* 2'-O-methyl	Wait Time* Ribo
Phosphoramidites	22/33/66	40/60/120 µL	60 sec	180 sec	360sec
S-Ethyl Tetrazole	70/105/210	40/60/120 µL	60 sec	180 min	360 sec
Acetic Anhydride	265/265/265	50/50/50 µL	10 sec	10 sec	10 sec
N-Methyl Imidazole	502/502/502	50/50/50 µL	10 sec	10 sec	10 sec
TCA	238/475/475	250/500/500 µL	15 sec	15 sec	15 sec
Iodine	6.8/6.8/6.8	80/80/80 µL	30 sec	30 sec	30 sec
Beaucage	34/51/51	80/120/120	100 sec	200 sec	200 sec
Acetonitrile	NA	1150/1150/1150 µL	NA	NA	NA

• **Wait time does not include contact time during delivery.**

Table 3: Peptides for Conjugation

Peptide	Sequence	SEQ ID NO
ANTENNAPEDIA	RQI KIW FQN RRM KWK K amide	14
Kaposi fibroblast growth factor	AAV ALL PAV LLA LLA P + VQR KRQ KLMP	15
caiman crocodylus Ig(5) light chain	MGL GLH LLV LAA ALQ GA	16
HIV envelope glycoprotein gp41	GAL FLG FLG AAG STM GA + PKS KRK 5 (NLS of the SV40)	17
HIV-1 Tat	RKK RRQ RRR	18
Influenza hemagglutinin envelop glycoprotein	GLFEAIAGFIENGWEGMIDGGGYC	19
RGD peptide	X-RGD-X where X is any amino acid or peptide	20
transportan A	GWT LNS AGY LLG KIN LKA LAA LAK KIL	21
Somatostatin (tyr-3-octreotate)	(S)*FC* Y*WK* TCT	22
Pre-S-peptide	(S)DH QLN PAF	23

(S) optional Serine for coupling

Italic = optional D isomer for stability

Claims

1.   A compound having the Formula 1:
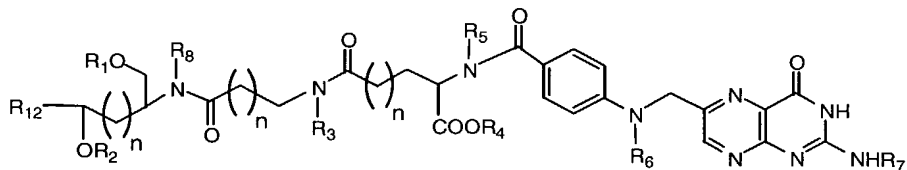
wherein each R_1, R_3, R_4, R_5, R_6, R_7 and R_8 is independently hydrogen, alkyl , substituted alkyl, aryl, substituted aryl, or a protecting group, each "n" is independently an integer from 0 to about 200, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and R_2 is a phosphorus containing group, nucleoside, nucleotide, small molecule, nucleic acid, or a solid support comprising a linker.

2.   A compound having the Formula 2:
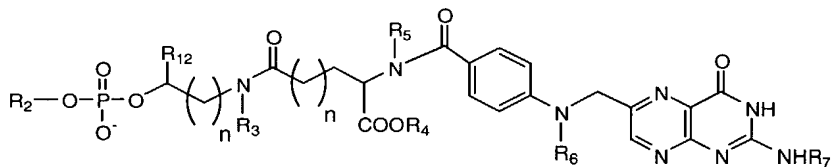
wherein each R_3, R_4, R_5, R_6 and R_7 is independently hydrogen, alkyl , substituted alkyl, aryl, substituted aryl, or a protecting group, each "n" is independently an integer from 0 to about 200, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and R_2 is a phosphorus containing group, nucleoside, nucleotide, small molecule, nucleic acid, or a solid support comprising a linker.

3.   A compound having the Formula 3:
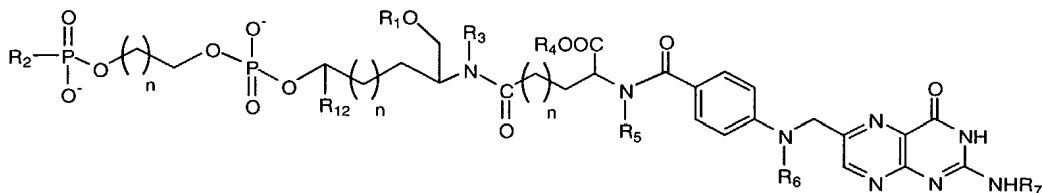
wherein each R_1, R_3, R_4, R_5 R_6 and R_7 is independently hydrogen, alkyl , substituted alkyl, aryl, substituted aryl, or a protecting group, each "n" is independently an integer from 0 to about 200, R_{12} is a straight or branched chain

alkyl, substituted alkyl, aryl, or substituted aryl, and R_2 is a phosphorus containing group, nucleoside, nucleotide, small molecule, or nucleic acid.

4.   A compound having the Formula 4:

wherein each R_3, R_4, R_5, R_6 and R_7 is independently hydrogen, alkyl, substituted alkyl, aryl, substituted aryl, or a protecting group, each "n" is independently an integer from 0 to about 200, R_2 is a phosphorus containing group, nucleoside, nucleotide, small molecule, nucleic acid, or a solid support comprising a linker, and R_{13} is an amino acid side chain.

5.   A compound having the Formula 5:

wherein each $R1$ and $R4$ is independently a protecting group or hydrogen, each R_3, R_5, R_6, R_7 and R_8 is independently hydrogen, alkyl or nitrogen protecting group, each "n" is independently an integer from 0 to about 200, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and each R_9 and R_{10} is independently a nitrogen containing group, cyanoalkoxy, alkoxy, aryloxy, or alkyl group.

6.   A compound having the Formula 6:
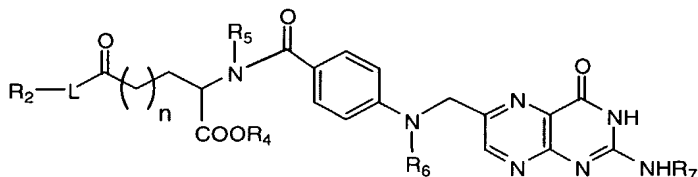
wherein each R_4, R_5, R_6 and R_7 is independently hydrogen, alkyl, substituted alkyl, aryl, substituted aryl, or a protecting group, R_2 is a phosphorus containing

group, nucleoside, nucleotide, small molecule, nucleic acid, or a solid support comprising a linker, each "n" is independently an integer from 0 to about 200, and L is a degradable linker.

7.   A compound having the Formula 7:
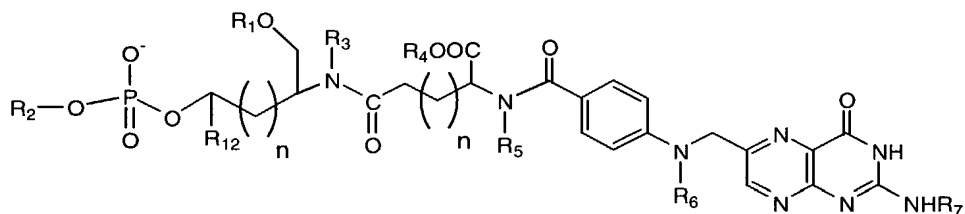
5

wherein each R_1, R_3, R_4, R_5, R_6 and R_7 is independently hydrogen, alkyl , substituted alkyl, aryl, substituted aryl, or a protecting group, each "n" is independently an integer from 0 to about 200, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and R_2 is a phosphorus containing

10   group, nucleoside, nucleotide, small molecule, nucleic acid, or a solid support comprising a linker.

8.   A compound having the Formula 8:
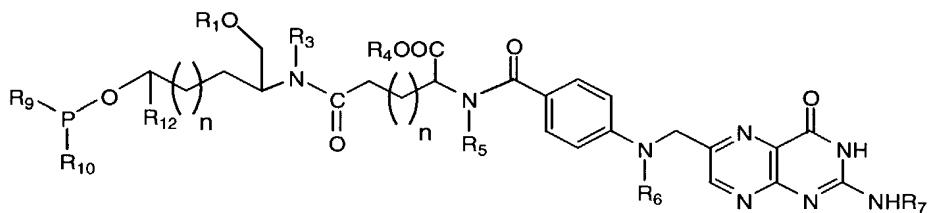
wherein each R1 and R4 is independently a protecting group or hydrogen, each

15   R_3, R_5, R_6 and R_7 is independently hydrogen, alkyl or nitrogen protecting group, each "n" is independently an integer from 0 to about 200, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and each R_9 and R_{10} is independently a nitrogen containing group, cyanoalkoxy, alkoxy, aryloxy, or alkyl group.

20

9.   A method for synthesizing a compound having Formula 5:

wherein each R1 and R4 is independently a protecting group or hydrogen, each R_3, R_5, R_6 and R_7 is independently hydrogen, alkyl or nitrogen protecting group, each "n" is independently an integer from 0 to about 200, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and each R_9 and R_{10} is independently a nitrogen containing group, cyanoalkoxy, alkoxy, aryloxy, or alkyl group, comprising;

(a) coupling a bis-hydroxy aminoalkyl derivative with a N-protected aminoalkanoic acid to yield a compound of Formula 9;
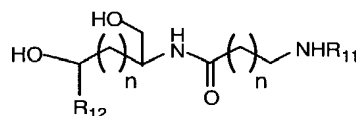
wherein R_{11} is an amino protecting group, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and each "n" is independently an integer from 0 to about 200;

(b) introducing primary hydroxy protection followed by amino deprotection to yield a compound of Formula 10;
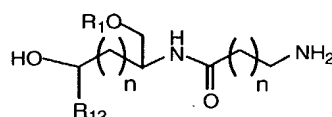
wherein R_1 is a protecting group, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and each "n" is independently an integer from 0 to about 200;

(c) coupling the deprotected amine with a protected amino acid to yield a compound of Formula 11;
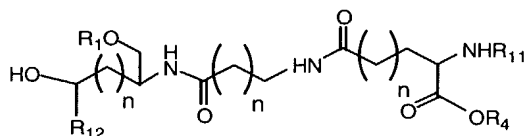
wherein each R1 and R4 is independently a protecting group or hydrogen, each "n" is independently an integer from 0 to about 200, R_{11} is an amino protecting group, and R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl;

(d) deprotecting the amine of the conjugated glutamic acid to yield a compound of Formula 12;
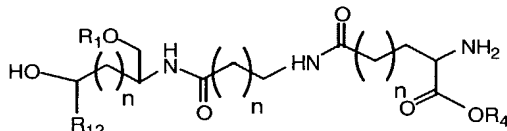
wherein each R1 and R4 is independently a protecting group or hydrogen, each "n" is independently an integer from 0 to about 200, R_{11} is an amino protecting group, and R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl;

(e) coupling the deprotected amine with an amino protected pteroic acid to yield a compound of Formula 13;

wherein each R_1 and R_4 is independently a protecting group or hydrogen, each R_3, R_5, R_6 and R_7 is independently hydrogen, alkyl or nitrogen protecting group, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and each "n" is independently an integer from 0 to about 200; and

(f) introducing a phosphorus containing group at the secondary hydroxyl to yield a compound of Formula 5.

10.   A method for synthesizing a compound having Formula 8:
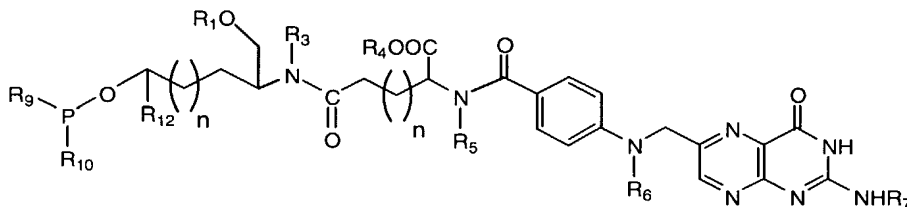
wherein each R_1 and R_4 is independently a protecting group or hydrogen, each R_3, R_5, R_6 and R_7 is independently hydrogen, alkyl or nitrogen protecting group, each

"n" is independently an integer from 0 to about 200, each R_9 and R_{10} is independently a nitrogen containing group, cyanoalkoxy, alkoxy, aryloxy, or alkyl group, and R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, comprising;

5   (a) coupling a bis-hydroxy aminoalkyl derivative with a protected amino acid to yield a compound of Formula 14;
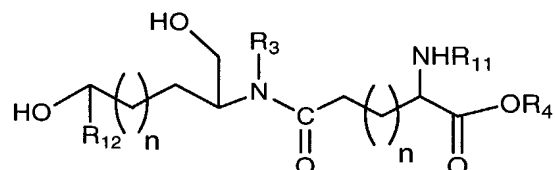
wherein R_{11} is an amino protecting group, each "n" is independently an integer from 0 to about 200, R_4 is independently a protecting group, and R_{12} is a straight

10   or branched chain alkyl, substituted alkyl, aryl, or substituted aryl;

(b) introducing primary hydroxy protection followed by amino deprotection to yield a compound of Formula 15;
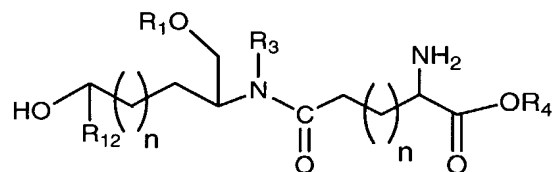
wherein each R_1 and R_4 is independently a protecting group or hydrogen, R_{12} is a

15   straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and each "n" is independently an integer from 0 to about 200;

(c) coupling the deprotected amine with an amino protected pteroic acid to yield a compound of Formula 16;
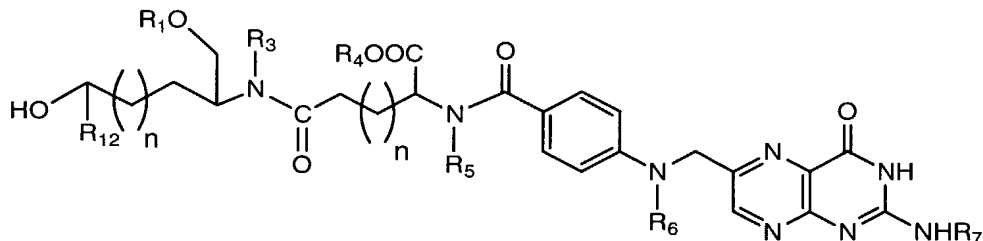
20   wherein each R_1 and R_4 is independently a protecting group or hydrogen, each R_3, R_5, R_6 and R_7 is independently hydrogen, alkyl or nitrogen protecting group, R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl, and each "n" is independently an integer from 0 to about 200; and

SUBSTITUTE SHEET (RULE 26)

(f) introducing a phosphorus containing group at the secondary hydroxyl to yield a compound of Formula 18.

11. The compound of any of claims 1, 2, 3, 4, 6, or 7, wherein R_2 is a phosphorus containing group.

12. The compound of any of claims 1, 2, 3, 4, 6, or 7, wherein R_2 is a nucleoside.

13. The compound of any of claims 1, 2, 3, 4, 6, or 7, wherein R_2 is a nucleotide.

14. The compound of any of claims 1, 2, 3, 4, 6, or 7, wherein R_2 is a small molecule.

15. The compound of any of claims 1, 2, 3, 4, 6, or 7, wherein R_2 is a nucleic acid.

16. The compound of any of claims 1, 2, 3, 4, 6, or 7, wherein R_2 is a solid support comprising a linker.

17. The compound of claim 12, wherein said nucleoside is a nucleoside with anticancer activity.

18. The compound of claim 12, wherein said nucleoside is a nucleoside with antiviral activity.

19. The compound of claim 12, wherein said nucleoside is fludarabine.

20. The compound of claim 12, wherein said nucleoside is lamivudine (3TC).

21. The compound of claim 12, wherein said nucleoside is 5-fluro uridine.

22. The compound of claim 12, wherein said nucleoside is AZT.

23. The compound of claim 12, wherein said nucleoside is ara-adenosine or ara-adenosine monophosphate.

24. The compound of claim 12, wherein said nucleoside is a dideoxy nucleoside analog.

25. The compound of claim 12, wherein said nucleoside is carbodeoxyguanosine.

26. The compound of claim 12, wherein said nucleoside is ribavirin.

27. The compound of claim 12, wherein said nucleoside is fialuridine.

28. The compound of claim 12, wherein said nucleoside is lobucavir.

29. The compound of claim 12, wherein said nucleoside is a pyrophosphate nucleoside analog.

30. The compound of claim 12, wherein said nucleoside is an acyclic nucleoside analog.

31. The compound of claim 12, wherein said nucleoside is acyclovir.

32.   The compound of claim 12, wherein said nucleoside is gangciclovir.

33.   The compound of claim 12, wherein said nucleoside is penciclovir.

34.   The compound of claim 12, wherein said nucleoside is famciclovir.

35.   The compound of claim 12, wherein said nucleoside is an L-nucleoside analog.

36.   The compound of claim 12, wherein said nucleoside is FTC.

37.   The compound of claim 12, wherein said nucleoside is L-FMAU.

38.   The compound of claim 12, wherein said nucleoside is L-ddC or L-FddC.

39.   The compound of claim 12, wherein said nucleoside is L-d4C or L-Fd4C.

40.   The compound of claim 12, wherein said nucleoside is an L-dideoxypurine nucleoside analog.

41.   The compound of claim 12, wherein said nucleoside is cytallene.

42.   The compound of claim 12, wherein said nucleoside is bis-POM PMEA (GS-840).

43.   The compound of claim 12, wherein said nucleoside is BMS-200,475.

44.   The compound of claim 4, wherein R_{13} comprises an alkylamine.

45.   The compound of claim 4, wherein R_{13} comprises an alkanol.

46.   The compound of claim 4, wherein R_{13} comprises -CH_2O-.

47.   The compound of claim 4, wherein R_{13} comprises –$CH(CH_2)CH_2O$-.

48.   The compound of claim 6, wherein L is serine.

49.   The compound of claim 6, wherein L is threonine.

50.   The compound of claim 6, wherein L is a photolabile linkage.

51.   The compound of any of claims 5, 8, 9, or 10, wherein R_9 comprises a phosphorus protecting group

52.   The compound of claim 51, wherein said phosphorus protecting group is -OCH_2CH_2CN (oxyethylcyano).

53.   The compound of any of claims 5 or 8, wherein R_{10} comprises a nitrogen containing group.

54.   The compound of claim 53, wherein said nitrogen containing group is –$N(R_{14})$ wherein R_{14} is a straight or branched chain alkyl having form about 1 to 10 carbons.

55. The compound of any of claims 5 or 8, wherein R_{10} comprises a heterocycloalkyl or heterocycloalkenyl ring containing from about 4 to 7 atoms, and having up to 3 heteroatoms selected from oxygen, nitrogen, and sulfur.

56. The compound of any of claims 1, 5 or 8, wherein R_1 is an acid labile protecting group.

57. The compound of any of claims 1, 5 or 8, wherein R_1 is a trityl or substituted trityl group.

58. The compound of claim 57, wherein said substituted trityl group is a dimethoxytrityl or mono-methoxytrityl group.

59. The compound of any of claims 1, 2, 3, 4, 5, 6, 7 or 8, wherein R_4 is *tert*-butyl, Fm (fluorenyl-methoxy), or allyl.

60. The compound of any of claims 1, 2, 3, 4, 5, 6, 7 or 8, wherein R_6 is TFA (trifluoracetyl).

61. The compound of any of claims 1, 2, 3, 4, 5, 6, 7 or 8, wherein R_3, R_5 R_7 and R_8 are hydrogen.

62. The compound of any of claims 1, 2, 3, 4, 5, 6, 7 or 8, wherein R_7 is isobutyryl.

63. The compound of any of claims 1, 2, 3, 4, 5, 6, 7 or 8, wherein R_7 is dimethylformamide.

64. The compound of any of claims 1, 2, 3, 4, 5, 6, 7 or 8, wherein R_7 is hydrogen.

65. The compound of any of claims 1, 2, 3, 5, 7 or 8, wherein R_{12} is methyl.

66. The compound of any of claims 1, 2, 3, 5, 7 or 8, wherein R_{12} is ethyl.

67. The compound of any of claim 15, wherein said nucleic acid is an enzymatic nucleic acid.

68. The compound of claim 67, wherein said enzymatic nucleic acid is a hammerhead.

69. The compound of claim 67, wherein said enzymatic nucleic acid is an Inozyme.

70. The compound of claim 67, wherein said enzymatic nucleic acid is a DNAzyme.

71. The compound of claim 67, wherein said enzymatic nucleic acid is a G-cleaver.

72. The compound of claim 67, wherein said enzymatic nucleic acid is a Zinzyme.

73. The compound of claim 67, wherein said enzymatic nucleic acid is an Amberzyme.

74. The compound of claim 67, wherein said enzymatic nucleic acid is an allozyme.

75. The compound of any of claim 15, wherein said nucleic acid is an antisense nucleic acid.

76. The compound of any of claim 15, wherein said nucleic acid is a 2-5A nucleic acid chimera.

77. The compound of any of claim 15, wherein said nucleic acid is a decoy nucleic acid.

78. The compound of claim 13, wherein said nucleotide is a nucleotide with anticancer activity.

79. The compound of claim 13, wherein said nucleotide is a nucleotide with antiviral activity.

80. The compound of claim 16, wherein said solid support comprising a linker is of Formula 17:
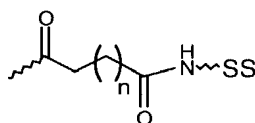
wherein SS is a solid support, and each "n" is independently an integer from 1 to 200.

81. The compound of claim 80, wherein said solid support is controlled pore glass (CPG).

82. The compound of claim 80, wherein said solid support is polystyrene.

83. The compound of claim 16, wherein said compound is used in the synthesis of a nucleic acid.

84. A pharmaceutical composition comprising the compound of claim 1 in a pharmaceutically acceptable carrier.

85. A pharmaceutical composition comprising the compound of claim 2 in a pharmaceutically acceptable carrier.

86. A pharmaceutical composition comprising the compound of claim 3 in a pharmaceutically acceptable carrier.

87. A pharmaceutical composition comprising the compound of claim 4 in a pharmaceutically acceptable carrier.

88. A pharmaceutical composition comprising the compound of claim 6 in a pharmaceutically acceptable carrier.

89.   A pharmaceutical composition comprising the compound of claim 7 in a pharmaceutically acceptable carrier.

90.   A method of treating cancer in a patient, comprising contacting cells of said patient with the pharmaceutical composition of any of claims 84-89, under conditions suitable for said treatment.

91.   The method of claim 90, further comprising the use of one or more other drug therapies under conditions suitable for said treatment.

92.   The method of claim 90, wherein said cancer is breast cancer, lung cancer, colorectal cancer, brain cancer, esophageal cancer, stomach cancer, bladder cancer, pancreatic cancer, cervical cancer, head and neck cancer, ovarian cancer, melanoma, lymphoma, glioma, or multidrug resistant cancers.

93.   A method of treating a patient infected with a virus, comprising contacting cells of said patient with the pharmaceutical composition of any of claims 84-89, under conditions suitable for said treatment.

94.   The method of claim 93, further comprising the use of one or more other drug therapies under conditions suitable for said treatment.

95.   The method of claim 93, wherein said virus is HIV, HBV, HCV, CMV, RSV, HSV, poliovirus, influenza, rhinovirus, west nile virus, Ebola virus, foot and mouth virus, and papilloma virus.

96.   A kit for detecting the presence of a nucleic acid in a sample, comprising the compound of claim 15.

97.   A kit for detecting the presence of a target molecule in a sample, comprising the compound of claim 15.

98.   A kit for detecting the presence of a nucleic acid in a cancer cell, comprising the compound of any of claim 74.

99.   A kit for detecting the presence of a nucleic acid in a virus infected cell, comprising the compound of claim 74.

100.   The compound of any of claims 2, 3, 4, or 7, wherein said compound contains a modified phosphate.

101.   The compound of any of claims 1, 2, 3, 4, 6, or 7, wherein said phosphorus containing group is a phosphoramidite, phosphodiester, phosphoramidate,

phosphorothioate, phosphorodithioate, alkylphosphonate, arylphosphonate, monophosphate, diphosphate, triphosphate, or pyrophosphate.

102. The compound of claim 12, wherein said nucleoside is carbovir or abacavir.

103. A method for synthesizing a compound having Formula 18:
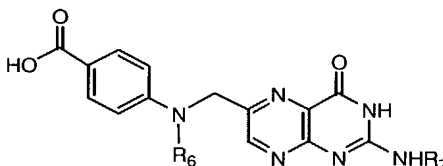
wherein each R_6 and R_7 is independently hydrogen, alkyl or nitrogen protecting group, comprising;

(a) treating folic acid with a carboxypeptidase to yield a compound of Formula 19;
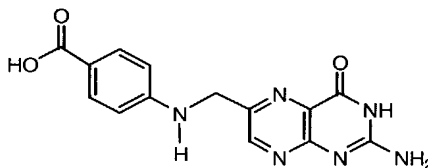
(b) introducing protection of the secondary amine to yield a compound of Formula 20;
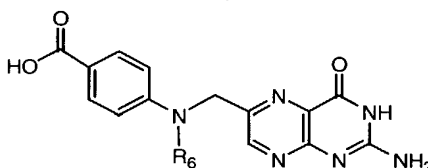
wherein R_6 is a nitrogen protecting group; and

(c) introducing protection of the primary amine to yield a compound of Formula 18.

104. The method of claim 103, wherein R_6 is trifluoroacetyl (TFA).

105. The method of claim 103, wherein R_7 is isobutyryl (iBu).

106. The method of claim 9, wherein said amino protected pteroic acid is a compound of Formula 18.

107. The method of claim 10, wherein said amino protected pteroic acid is a compound of Formula 18.

108. A compound of claim 1, having Formula 21:

146
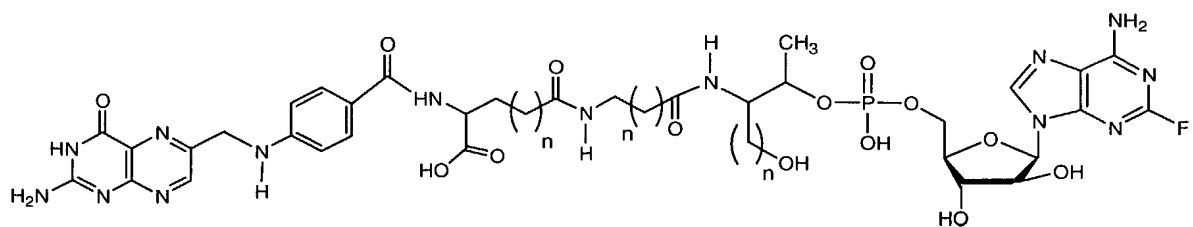
wherein each "n" is independently an integer from 0 to about 200.

109.   A compound of claim 4, having Formula 22:

5
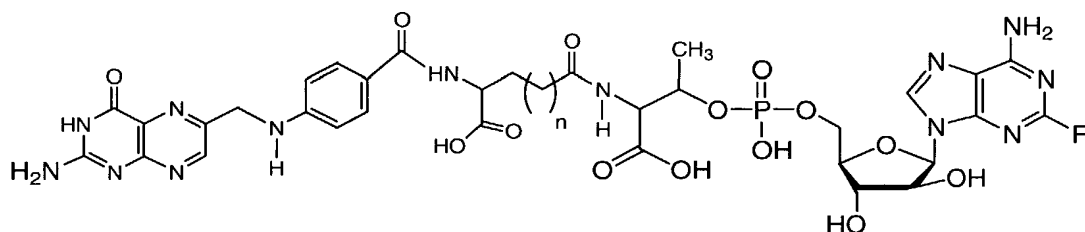
wherein each "n" is independently an integer from 0 to about 200.

110.   A compound of claim 7, having Formula 23:
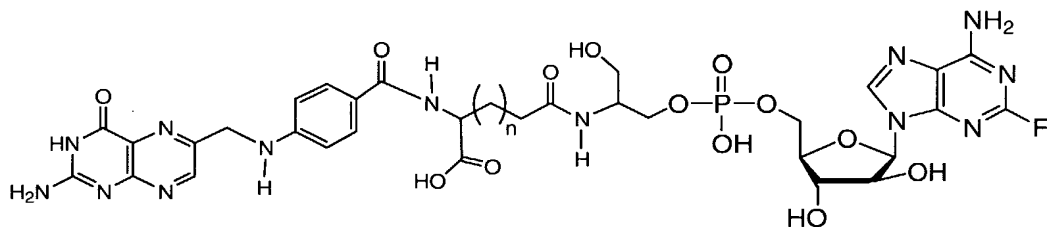
wherein "n" is an integer from 0 to about 200.

10   111.   A compound having Formula 24:
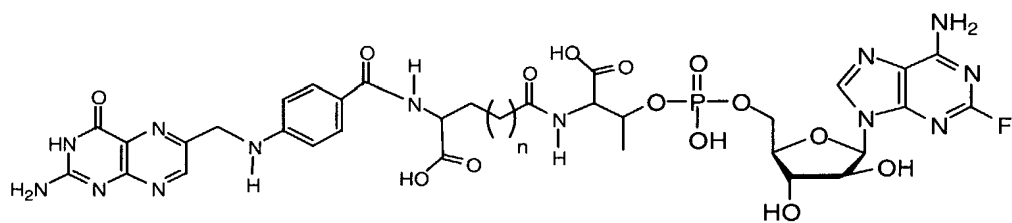
wherein "n" is an integer from 0 to about 200.

112.   A compound having Formula 25:
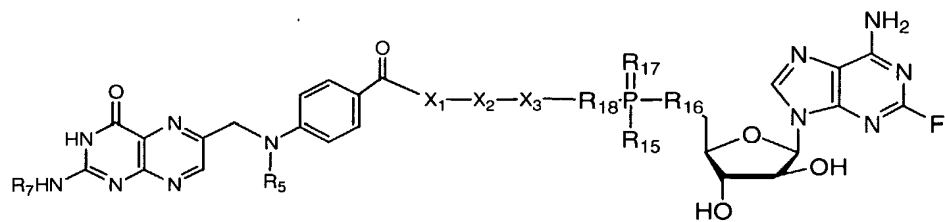
wherein each R_5 and R_7 is independently hydrogen, alkyl or a nitrogen protecting group, each R_{15}, R_{16}, R_{17}, and R_{18} is independently O, S, alkyl, substituted alkyl, aryl, substituted aryl, or halogen, X_1 is $-CH(X_{1'})$ or a group of Formula 38:
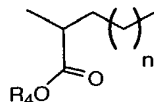
wherein R_4 is a protecting group and "n" is an integer from 0 to about 200;

$X_{1'}$ is the protected or unprotected side chain of a naturally occurring or non-naturally-occurring amino acid, X_2 is an amide, alkyl, or carbonyl containing linker or a bond, and X_3 is a degradable linker which is optionally absent.

113.   The compound of claim 112, wherein X_3 is a group of Formula 26:
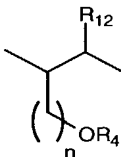
wherein R_4 is hydrogen or a protecting group, "n" is an integer from 0 to about 200 and R_{12} is a straight or branched chain alkyl, substituted alkyl, aryl, or substituted aryl.

114.   The compound of claim 113, wherein R_4 is hydrogen and R_{12} is methyl or hyrdogen.

115.   A pharmaceutical composition comprising the compound of claim 108 in a pharmaceutically acceptable carrier.

116.   A pharmaceutical composition comprising the compound of claim 109 in a pharmaceutically acceptable carrier.

117.   A pharmaceutical composition comprising the compound of claim 110 in a pharmaceutically acceptable carrier.

118.   A pharmaceutical composition comprising the compound of claim 111 in a pharmaceutically acceptable carrier.

119.   A pharmaceutical composition comprising the compound of claim 112 in a pharmaceutically acceptable carrier.

120. A method of treating a cancer patient, comprising contacting cells of said patient with the pharmaceutical composition of any of claims 115-119, under conditions suitable for said treatment.

121. The method of claim 120, further comprising the use of one or more other therapies under conditions suitable for said treatment.

122. The method of claim 120, wherein said cancer is breast cancer, lung cancer, colorectal cancer, brain cancer, esophageal cancer, stomach cancer, bladder cancer, pancreatic cancer, cervical cancer, head and neck cancer, ovarian cancer, melanoma, lymphoma, glioma, or multidrug resistant cancers.

123. The compound of claim 2, wherein R_{12} is a alkylhyrdroxyl.

124. The compound of claim 123, wherein said alkylhydroxyl is $-(CH_2)_nOH$.

125. The compound of claim 124, wherein said n is an integer from 1-10.

126. The kit of claim 96, wherein said sample is from a cancer cell.

127. The kit of claim 96, wherein said sample is from a virus infected cell.

128. A compound having Formula 27:
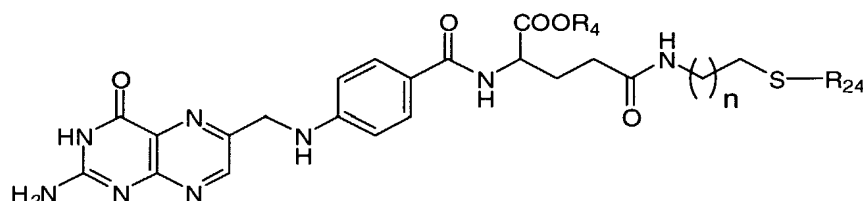
wherein "n" is an integer from about 0 to about 20 and R_4 is H or a cationic salt.

129. A method for synthesizing a compound having Formula 27 comprising:

(a) Selective tritylation of the thiol of cysteamine under conditions suitable to yield a compound having Formula 28:

wherein "n" is an integer from about 0 to about 20 and R_{19} is a thiol protecting group;

(b) peptide coupling of the product of (a) with a compound having Formula 29:

$$R_{21}HN\diagdown\diagup\diagdown COOH$$
$$COOR_{20}$$

wherein R_{20} is a carboxylic acid protecting group and R_{21} is an amino protecting group, under conditions suitable to yield a compound having Formula 30:

$$R_{21}HN\diagdown\diagup\diagdown\underset{\underset{COOR_{20}}{|}}{C}\underset{O}{\overset{O}{\parallel}}\underset{H}{N}()_n\diagdown SR_{19}$$

5   wherein "n" is an integer from about 0 to about 20, R_{19} is a thiol protecting group, R_{20} is a carboxylic acid protecting group and R_{21} is an amino protecting group;

(c) removing the amino protecting group R_{21} of the product of (b) under conditions suitable to yield a compound having Formula 31:

$$H_2N\diagdown\diagup\diagdown\underset{\underset{COOR_{20}}{|}}{C}\underset{H}{\overset{O}{\parallel}}\underset{}{N}()_n\diagdown SR_{19}$$

10   wherein "n" is an integer from about 0 to about 20 and R_{19} and R_{20} are as described in (b);

(d) condensation of the product of (c) with a compound having Formula 32:

$$R_{22}HN\diagtri COOH$$

wherein R_{22} is an amino protecting group, under conditions suitable to yield a

15   compound having Formula 33:

$$R_{22}HN\diagtri\underset{O}{\overset{O}{\parallel}}\underset{H}{N}\underset{\overset{|}{COOR_{20}}}{\diagup}\diagdown\underset{O}{\overset{}{\diagdown}}\underset{H}{N}()_n\diagdown SR_{19}$$

wherein "n" is an integer from about 0 to about 20 and R_{19} and R_{20} are as described in (b) and R_{22} is as described in (d);

(e) selective cleave of R_{22} from the product of (d) under conditions suitable to yield a compound having Formula 34:
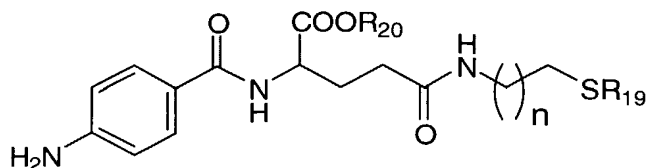
wherein "n" is an integer from about 0 to about 20 and R_{19} and R_{20} are as described in (b);

(f) coupling the product of (e) with a compound having Formula 35:
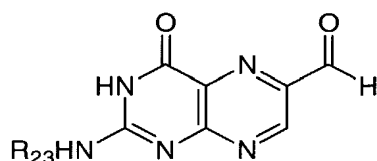
wherein R_{23} is an amino protecting group under conditions suitable to yield a compound having Formula 36:
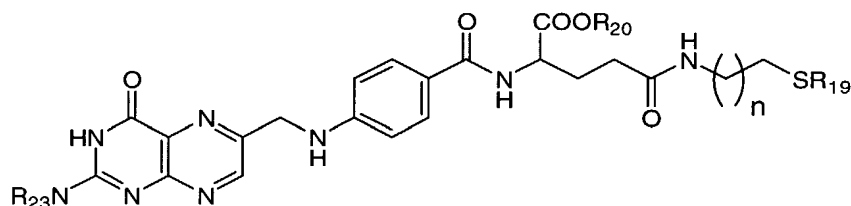
wherein R_{23} is an amino protecting group, "n" is an integer from about 0 to about 20 and R_{19} and R_{20} are as described in (b);

(g) deprotecting the product of (f) under conditions suitable to yield a compound having Formula 37.
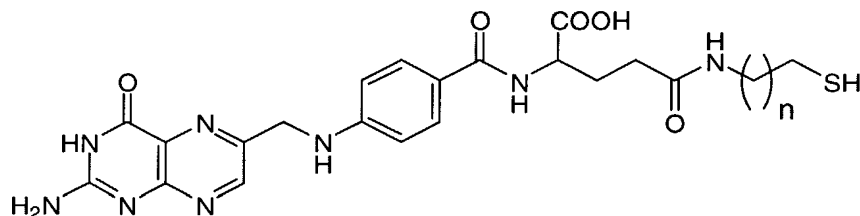
wherein "n" is an integer from about 0 to about 20; and

(h) introducing a disulphide-based leaving group to the product of (g) under conditions suitable to yield a compound having Formula 27.

130.   A compound having Formula 39:
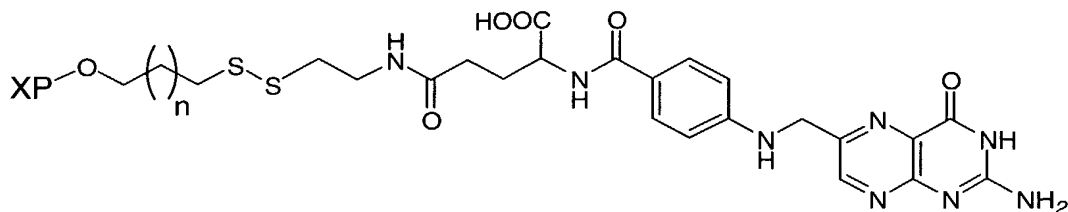
wherein "n" is an integer from about 0 to about 20, X is a nucleic acid, polynucleotide, or oligonucleotide, and P is a phosphorus containing group.

5   131.   A method for synthesizing a compound having Formula 39, comprising:

(a) Coupling a thiol containing linker to a nucleic acid, polynucleotide or oligonucleotide under conditions suitable to yield a compound having Formula 40:
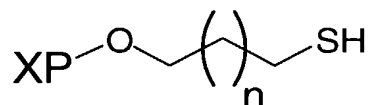
10   wherein "n" is an integer from about 0 to about 20, X is a nucleic acid, polynucleotide, or oligonucleotide, and P is a phosphorus containing group; and

(b) coupling the product of (a) with a compound having Formula 27 under conditions suitable to yield a compound having Formula 39.

132.   The method of claim 131, wherein said thiol containing linker is a compound
15   having Formula 41:
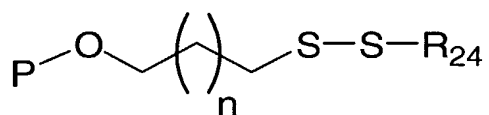
wherein "n" is an integer from about 0 to about 20, P is a phosphorus containing group, and R24 is any alkyl, substituted alkyl, alkoxy, aryl, substituted aryl, alkenyl, substituted alkenyl, alkynyl, or substituted alkynyl group with or without
20   additional protecting groups.

133.   The method of claim 131, wherein said conditions suitable to yield a compound having Formula 40 comprises reduction of the disulfide bond of a compound having Formula 42:
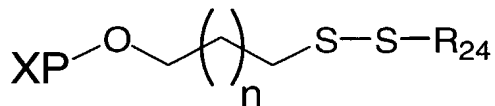
$$XP-O-\underbrace{}_{n}-S-S-R_{24}$$

wherein "n" is an integer from about 0 to about 20, X is a nucleic acid, polynucleotide, or oligonucleotide, P is a phosphorus containing group, and R_{24} is any alkyl, substituted alkyl, alkoxy, aryl, substituted aryl, alkenyl, substituted alkenyl, alkynyl, or substituted alkynyl group with or without additional protecting groups.

134.   A pharmaceutical composition comprising the compound of claim 128 in a pharmaceutically acceptable carrier.

135.   A pharmaceutical composition comprising the compound of claim 130 in a pharmaceutically acceptable carrier.

136.   A method of treating a cancer patient, comprising contacting cells of said patient with the pharmaceutical composition of any claim 134 or claim 135, under conditions suitable for said treatment.

137.   A compound having Formula 43:
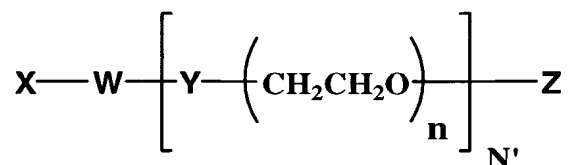
$$X-W-\left[Y-\left(CH_2CH_2O\right)_n\right]_{N'}-Z$$

43

wherein X comprises a biologically active molecule; W comprises a degradable nucleic acid linker; Y comprises a linker molecule or amino acid that can be present or absent; Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; n is an integer from about 1 to about 100; and N' is an integer from about 1 to about 20.

138.   A compound having Formula 44:
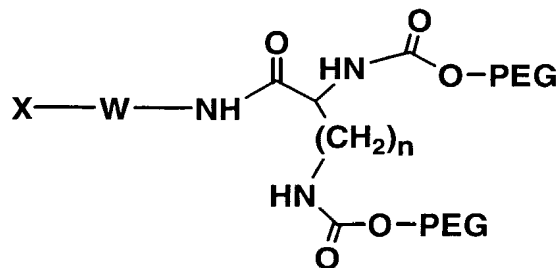
$$\text{X—W—NH—C(=O)—CH(—HN—C(=O)—O–PEG)—(CH}_2)_n\text{—HN—C(=O)—O–PEG}$$

44

wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent; n is an integer from about 1 to about 50, and PEG represents a compound having Formula 45:

$$-\left[\text{CH}_2\text{CH}_2\text{O}\right]_n\text{—Z}$$

45
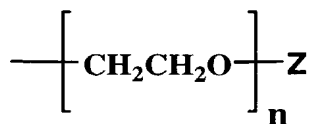
wherein Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; and n is an integer from about 1 to about 100.

139.   A compound having Formula 46:
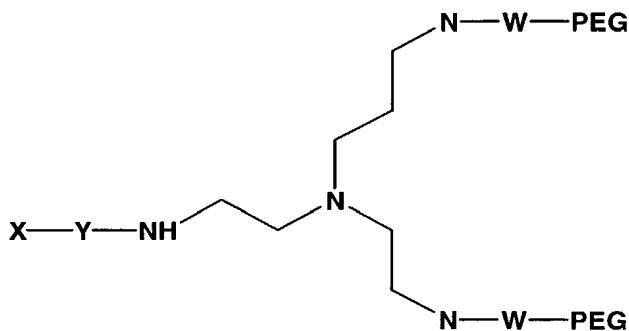
$$\text{X—Y—NH—CH}_2\text{CH}_2\text{—N}\left(\text{CH}_2\text{CH}_2\text{CH}_2\text{—N—W—PEG}\right)\left(\text{CH}_2\text{CH}_2\text{—N—W—PEG}\right)$$

46

wherein X comprises a biologically active molecule; each W independently comprises linker molecule or chemical linkage that can be present or absent, Y comprises a linker molecule or chemical linkage that

can be present or absent; and PEG represents a compound having Formula 45:

$$-\left[CH_2CH_2O\right]_n Z$$

45
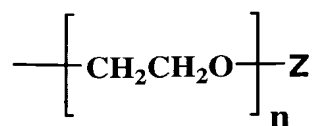
wherein Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; and n is an integer from about 1 to about 100.

140.   A compound having Formula 47:

$$X-W-Y-R_1-\overset{\overset{R_4}{\|}}{\underset{\underset{R_2}{|}}{P}}-R_3-W-\left(Q\right)_n$$

47

wherein X comprises a biologically active molecule; each W independently comprises a linker molecule or chemical linkage that can be the same or different and can be present or absent, Y comprises a linker molecule that can be present or absent; each Q independently comprises a hydrophobic group or phospholipid; each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, and n is an integer from about 1 to about 10.

141.   A compound having Formula 48:
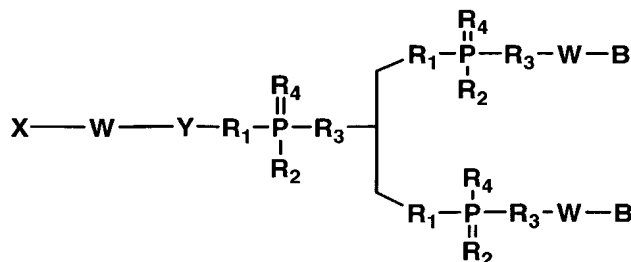
$$X-W-Y-R_1-\overset{\overset{R_4}{\|}}{\underset{\underset{R_2}{|}}{P}}-R_3\left\{\begin{array}{l} R_1-\overset{\overset{R_4}{\|}}{\underset{\underset{R_2}{|}}{P}}-R_3-W-B \\ \\ R_1-\overset{\underset{R_4}{|}}{\underset{\overset{R_2}{\|}}{P}}-R_3-W-B \end{array}\right.$$

48

wherein X comprises a biologically active molecule; each W
independently comprises a linker molecule or chemical linkage that can be
present or absent, Y comprises a linker molecule that can be present or
absent; each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl,
alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted
N, and B represents a lipophilic group.

142.   A compound having Formula 49:
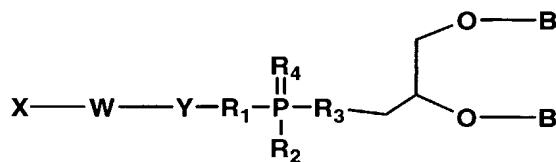
49

wherein X comprises a biologically active molecule; W comprises a linker
molecule or chemical linkage that can be present or absent, Y comprises a
linker molecule that can be present or absent; each R1, R2, R3, and R4
independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-
alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, and B represents
a lipophilic group.

143.   A compound having Formula 50:

50
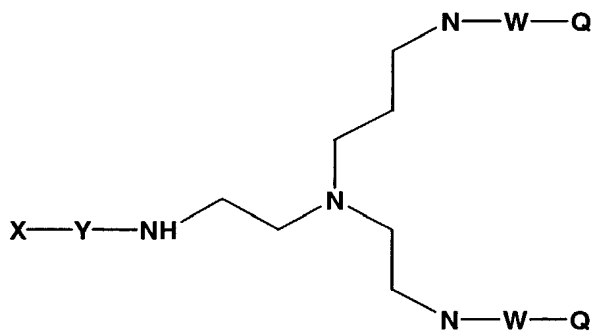
wherein X comprises a biologically active molecule; W comprises linker
molecule or chemical linkage that can be present or absent, Y comprises a
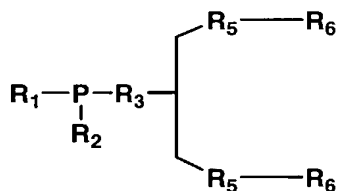linker molecule or chemical linkage that can be present or absent; and each
Q independently comprises a hydrophobic group or phospholipid.
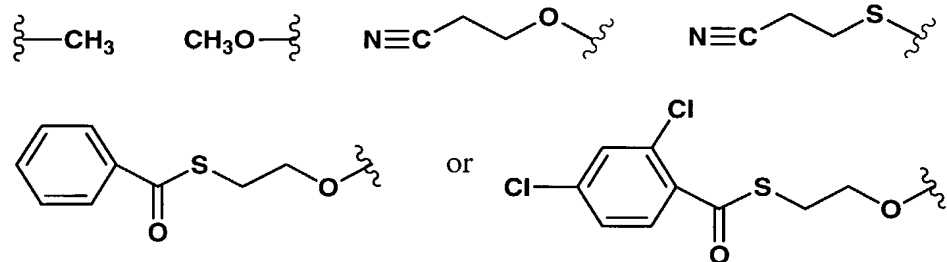
144.   A compound having Formula 51:
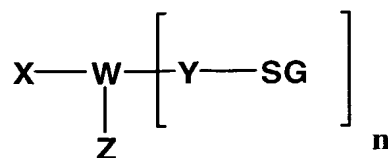
$$X-W-[Y-SG]_n$$
$$|$$
$$Z$$

51

wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent; Y comprises a linker molecule or amino acid that can be present or absent; Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; SG comprises a sugar and n is an integer from about 1 to about 20.

145.   A compound having Formula 52:
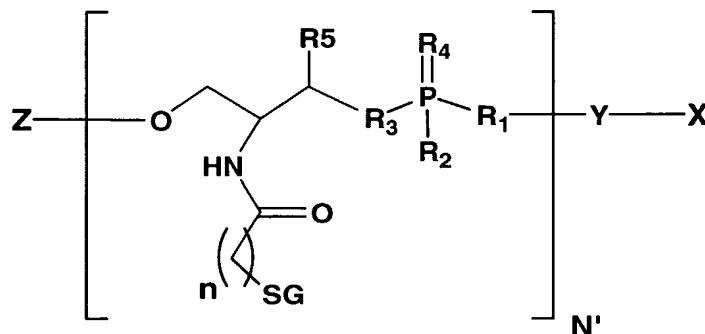
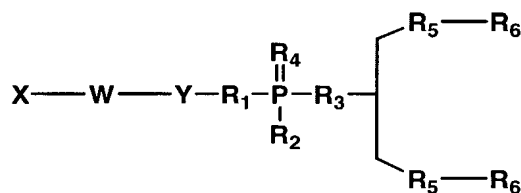52

wherein X comprises a biologically active molecule; Y comprises a linker molecule or chemical linkage that can be present or absent; each R1, R2, R3, R4, and R5 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N; Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; SG comprises a sugar, n is an integer from about 1 to about 20; and N' is an integer from about 1 to about 20.

146.   A compound having Formula 53:

157

53

wherein B comprises H, a nucleoside base, or a non-nucleosidic base with or without protecting groups; each R1 independently comprises O, N, S, alkyl, or substituted N; each R2 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylhalo, S, N, substituted N, or a phosphorus containing group; each R3 independently comprises N or O-N, each R4 independently comprises O, CH2, S, sulfone, or sulfoxy; X comprises H, a removable protecting group, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, or label; W comprises a linker molecule or chemical linkage that can be present or absent; SG comprises a sugar, each n is independently an integer from about 1 to about 50; and N' is an integer from about 1 to about 10.

147.   A compound having Formula 54:
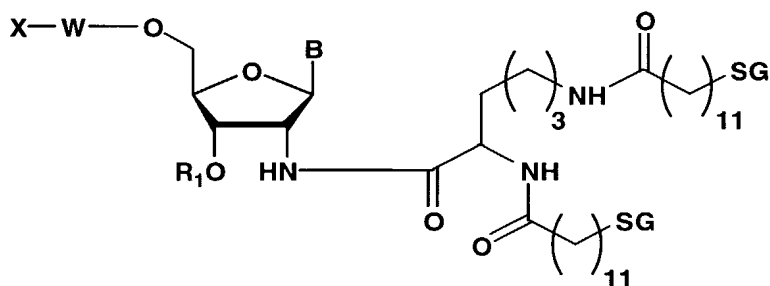
54

wherein B comprises H, a nucleoside base, or a non-nucleosidic base with or without protecting groups; each R1 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylhalo, S, N, substituted N, or a phosphorus containing group; X comprises H, a removable protecting group, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide,

enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, or label; W comprises a linker molecule or chemical linkage that can be present or absent; and SG comprises a sugar.

148.   A compound having Formula 55:

55

wherein each R1 independently comprises O, N, S, alkyl, or substituted N; each R2 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylhalo, S, N, substituted N, or a phosphorus containing group; each R3 independently comprises H, OH, alkyl, substituted alkyl, or halo; X comprises H, a removable protecting group, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide,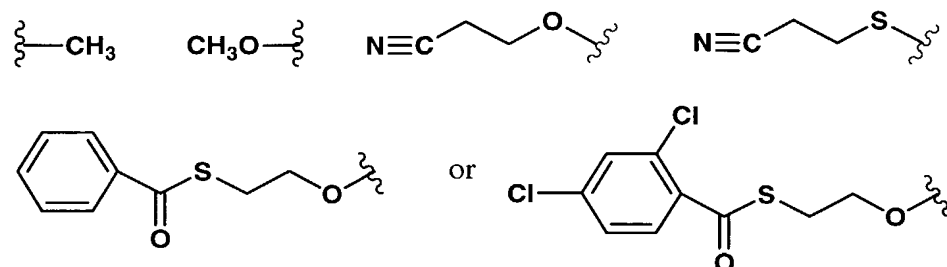 enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, biologically active molecule or label; W comprises a linker molecule or chemical linkage that can be present or absent; SG comprises a sugar, each n is independently an integer from about 1 to about 50; and N' is an integer from about 1 to about 100.

149.   A compound having Formula 56:

56

wherein R1 comprises H, alkyl, alkylhalo, N, substituted N, or a phosphorus containing group; R2 comprises H, O, OH, alkyl, alkylhalo, halo, S, N, substituted N, or a phosphorus containing group; X comprises H, a removable protecting group, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, biologically active molecule or label; W comprises a linker molecule or chemical linkage that can be present or absent; SG comprises a sugar, and each n is independently an integer from about 0 to about 20.

150.   A compound having Formula 57:
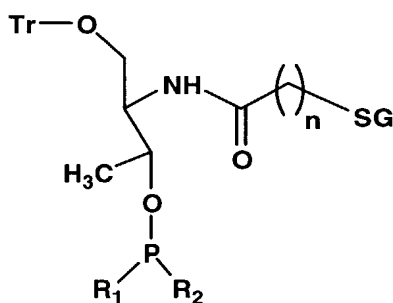
57

wherein R1 can include the groups:

and wherein R2 can include the groups:

and wherein Tr is a removable protecting group, for example a trityl, monomethoxytrityl, or dimethoxytrityl; SG comprises a sugar, and n is an integer from about 1 to about 20.

151.   A compound having Formula 58:
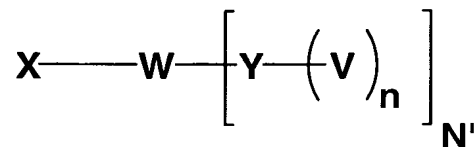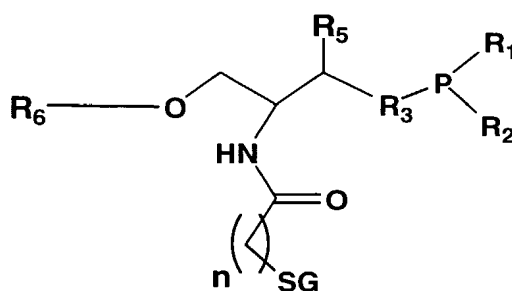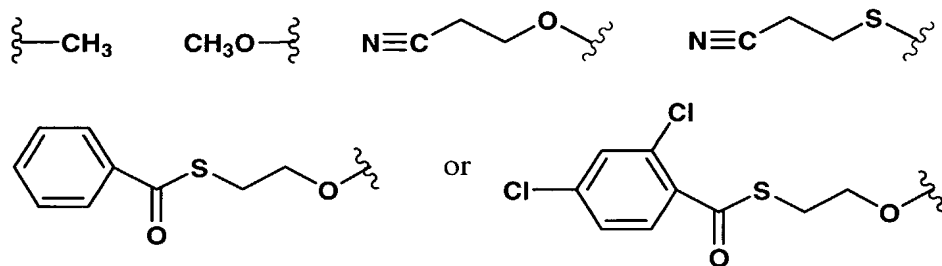
$$X\text{———}W\text{—}\left[Y\text{—}(V)_n\right]_{N'}$$

58

wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent; Y comprises a linker molecule or amino acid that can be present or absent; V comprises a protein or peptide; each n is independently an integer from about 1 to about 50; and N' is an integer from about 1 to about 100.

152.   A compound having Formula 59:

$$Q\text{-}R_2\text{---}\underset{\underset{\underset{X}{|}}{\underset{W}{|}}{\overset{R_1}{C}}\text{---}\underset{H}{N}\text{---}\overset{O}{C}\text{---}(CH_2)_n\text{---}O\text{—}N\text{—}W\text{———}V$$

59

wherein each R1 independently comprises O, S, N, substituted N, or a phosphorus containing group; each R2 independently comprises O, S, or N; X comprises H, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, or enzymatic nucleic acid or other biologically active molecule; n is an integer from about 1 to about 50, Q comprises H or a removable protecting group which can be optionally absent, each W independently comprises a linker molecule or chemical linkage that can be present or absent, and V comprises a protein or peptide or a compound having Formula 45

$$\text{—}\left[CH_2CH_2O\right]_n\text{-}Z$$

45

wherein Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; and n is an integer from about 1 to about 100.

5   153.   A compound having Formula 60:

60
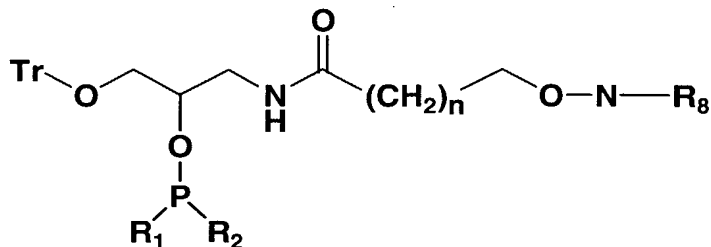
wherein R1 can include the groups:
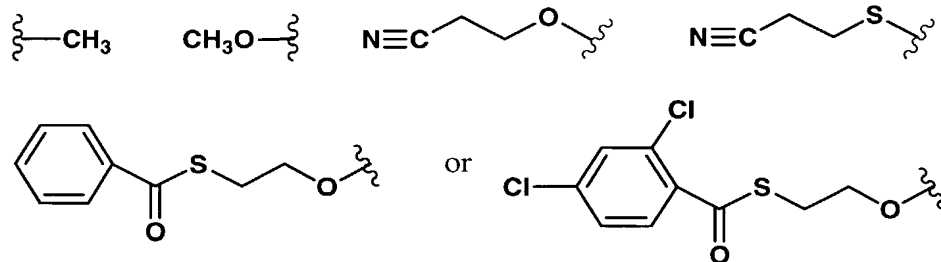
10   and wherein R2 can include the groups:
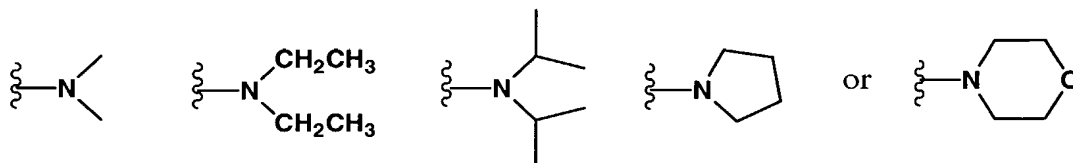
and wherein Tr is a removable protecting group, for example a trityl, monomethoxytrityl, or dimethoxytrityl; n is an integer from about 1 to about 50; and R8 is a nitrogen protecting group.

15   154.   A compound having Formula 61:
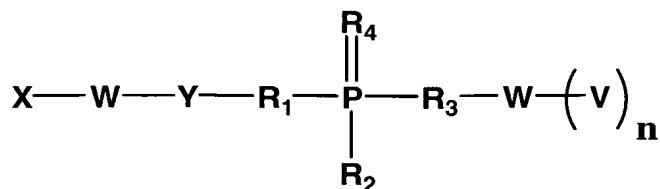
61

wherein X comprises a biologically active molecule; each W independently comprises a linker molecule or chemical linkage that can be the same or different and can be present or absent, Y comprises a linker molecule that can be present or absent; each 5 independently comprises a protein or peptide; each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, and n is an integer from about 1 to about 10.

155.   A compound having Formula 62:
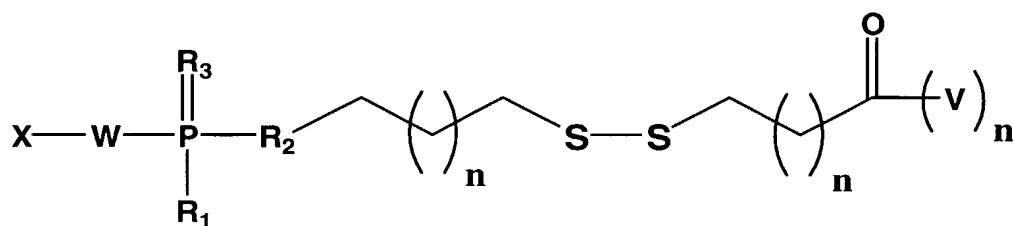
62

wherein X comprises a biologically active molecule; each 5 independently comprises a protein or peptide; W comprises a linker molecule or chemical linkage that can be present or absent; each R1, R2, and R3 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, and each n is independently an integer from about 1 to about 10.

156.   A compound having Formula 63:
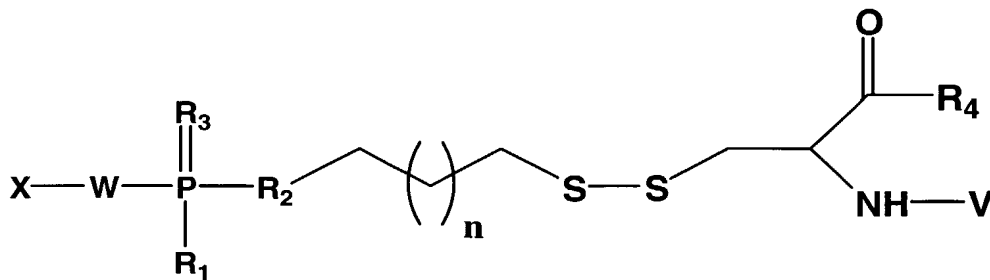
63

wherein X comprises a biologically active molecule; V comprises a protein or peptide; W comprises a linker molecule or chemical linkage that can be

present or absent; each R1, R2, R3 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, R4 represents an ester, amide, or protecting group, and each n is independently an integer from about 1 to about 10.

5   157.   A compound having Formula 64:
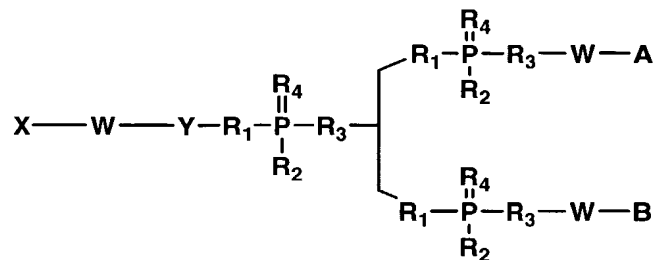
64

wherein X comprises a biologically active molecule; each W independently comprises a linker molecule or chemical linkage that can be present or absent, Y comprises a linker molecule that can be present or absent; each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, A comprises a nitrogen containing group, and B comprises a lipophilic group.

15   158.   A compound having Formula 65:
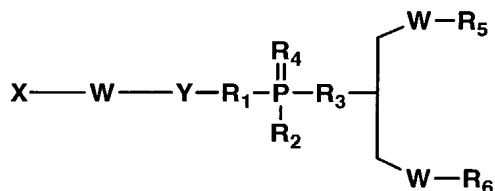
65

wherein X comprises a biologically active molecule; each W independently comprises a linker molecule or chemical linkage that can be present or absent, Y comprises a linker molecule that can be present or absent; each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, RV comprises the lipid or phospholipid component of any of Formulae 47-50, and R6 comprises a nitrogen containing group.

159.   In another embodiment, the invention features a compound having Formula 92:

92

wherein B comprises H, a nucleoside base, or a non-nucleosidic base with or without protecting groups; each R1 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylhalo, S, N, substituted N, or a phosphorus containing group; X comprises H, a removable protecting group, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, biologically active molecule or label; W comprises a linker molecule or chemical linkage that can be present or absent; R2 comprises O, NH, S, CO, COO, ON=C, or alkyl; R3 comprises alkyl, akloxy, or an aminoacyl side chain; and SG comprises a sugar.

160.   A compound having Formula 66:

66

wherein R1 comprises H, alkyl, alkylhalo, N, substituted N, or a phosphorus containing group; R2 comprises H, O, OH, alkyl, alkylhalo, halo, S, N, substituted N, or a phosphorus containing group; X comprises H, a removable protecting group, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, biologically active molecule or label; W comprises a linker molecule or chemical linkage that can be present or absent; R3 comprises O, NH, S, CO, COO, ON=C, or alkyl; R4

SUBSTITUTE SHEET (RULE 26)

comprises alkyl, akloxy, or an aminoacyl side chain; and SG comprises a sugar, for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose, fructose, or fucose and the respective D or L, alpha or beta isomers, and each n is independently an integer from about 0 to about 20.

161. A compound having Formula 87:

$$Y-W-\underset{\underset{R_1}{|}}{C}\!\equiv\!N-O-X$$

87

wherein X comprises a protein, peptide, antibody, lipid, phospholipid, oligosaccharide, label, biologically active molecule, for example a vitamin such as folate, vitamin A, E, B6, B12, coenzyme, antibiotic, antiviral, nucleic acid, nucleotide, nucleoside, or oligonucleotide such as an enzymatic nucleic acid, allozyme, antisense nucleic acid, siRNA, 2,5-A chimera, decoy, aptamer or triplex forming oligonucleotide, or polymers such as polyethylene glycol; W comprises a linker molecule or chemical linkage that can be present or absent; and Y comprises a biologically active molecule, for example an enzymatic nucleic acid, allozyme, antisense nucleic acid, siRNA, 2,5-A chimera, decoy, aptamer or triplex forming oligonucleotide, peptide, protein, or antibody; R1 comprises H, alkyl, or substituted alkyl.

162. A compound having Formula 88:
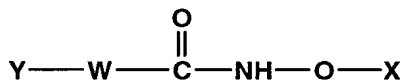
$$Y-W-\overset{\overset{\displaystyle O}{\displaystyle \|}}{C}-NH-O-X$$

88

wherein X comprises a protein, peptide, antibody, lipid, phospholipid, oligosaccharide, label, biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent, and Y comprises a biologically active molecule.

163. A method for the synthesis of a compound having Formula 48:

$$R_1-\overset{\overset{\displaystyle R_4}{\|}}{\underset{\underset{\displaystyle R_2}{|}}{P}}-R_3-W-B$$

$$X-W-Y-R_1-\overset{\overset{\displaystyle R_4}{\|}}{\underset{\underset{\displaystyle R_2}{|}}{P}}-R_3$$

$$R_1-\overset{\overset{\displaystyle R_4}{|}}{\underset{\underset{\displaystyle R_2}{\|}}{P}}-R_3-W-B$$

48

wherein X comprises a biologically active molecule; each W independently comprises a linker molecule or chemical linkage that can be present or absent, Y comprises a linker molecule that can be present or absent; each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N; and each B independently represents a lipophilic group, comprising: (a) introducing a compound having Formula 66:

$$R_1-\overset{}{\underset{\underset{\displaystyle R_2}{|}}{P}}-R_3-\begin{array}{l} R_5\text{——}R_6 \\ \\ R_5\text{——}R_6 \end{array}$$

66

wherein R1 is defined as in Formula 48 and can include the groups:

§—CH₃   CH₃O—§   N≡C~~~O~~~   N≡C~~~S~~~

(benzoyl-S-CH₂CH₂-O structure)   or   (dichlorobenzoyl-S-CH₂CH₂-O structure)

and wherein R2 is defined as in Formula 48 and can include the groups:

§—N(CH₃)₂   §—N(CH₂CH₃)₂   §—N(iPr)₂   §—N(pyrrolidine)   or   §—N(morpholine)

and wherein each R5 independently comprises O, N, or S and each R6 independently comprises a removable protecting group, for example a trityl, monomethoxytrityl, or dimethoxytrityl group, to a compound having Formula 67:

5

$$X \text{---} W \text{---} Y$$

67

wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent, and Y comprises a linker molecule that can be present or absent, under conditions

10   suitable for the formation of a compound having Formula 68:

68

wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent, Y comprises a

15   linker molecule that can be present or absent; and each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N comprising, each R5 independently comprises O, S, or N; and each R6 is independently a removable protecting group, for example a trityl, monomethoxytrityl, or

20   dimethoxytrityl group; (b) removing R6 from the compound having Formula 26 and (c) introducing a compound having Formula 69:

$$R_1 \text{---} \underset{\underset{R_2}{|}}{P} \text{---} R_3 \text{-} W \text{---} B$$

69

wherein R1 is defined as in Formula 48 and can include the groups:

and wherein R2 is defined as in Formula 48 and can include the groups:

or

and wherein W and B are defined as in Formula 48, to the compound having Formula 68 under conditions suitable for the formation of a compound having Formula 48.

164.   A method for the synthesis of a compound having Formula 49:
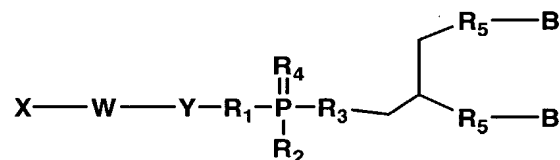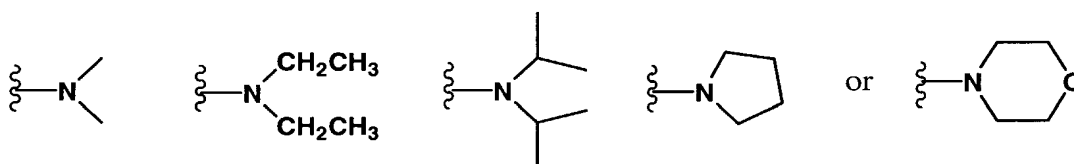
49

wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent, Y comprises a linker molecule that can be present or absent; each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N; each R5 independently comprises O, S, or N; and each B independently comprises a lipophilic group, comprising: (a) coupling a compound having Formula 70:

70
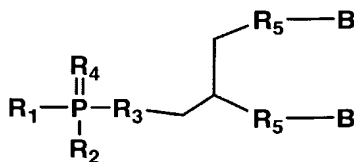
wherein R1 is defined as in Formula 49 and can include the groups:

and wherein R2 is defined as in Formula 49 and can include the groups:
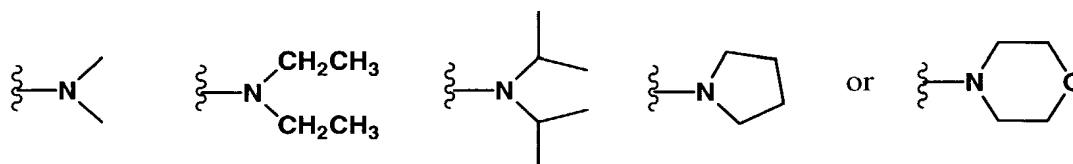
and wherein each R5 independently comprises O, S, or N, and wherein each B independently comprises a lipophilic group, with a compound having Formula 67:

$$X \longrightarrow W \longrightarrow Y$$

67

wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent, and Y comprises a linker molecule that can be present or absent, under conditions suitable for the formation of a compound having Formula 49.

165.   A method for the synthesis of a compound having Formula 52:
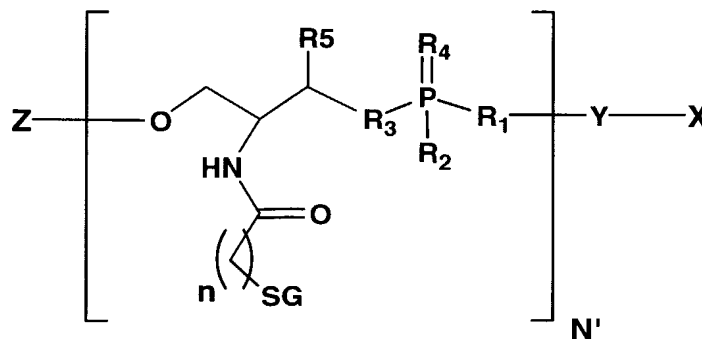
52

wherein X comprises a biologically active molecule; Y comprises a linker
molecule or chemical linkage that can be present or absent; each R1, R2,
R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl,
O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N; Z comprises H,
OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl,
amino, substituted amino, nucleotide, nucleoside, nucleic acid,
oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label;
SG comprises a sugar, n is an integer from about 1 to about 20; and N' is
an integer from about 1 to about 20, comprising: (a) coupling a compound
having Formula 71:

71

wherein R1, R2, R3, R5, SG, and n as is defined in Formula 10, and
wherein R1 can include the groups:

and wherein R2 can include the groups:

and R6 comprises a removable protecting group, for example a trityl,
monomethoxytrityl, or dimethoxytrityl group; with a compound having
Formula 72:

$$X \text{——} Y$$

72

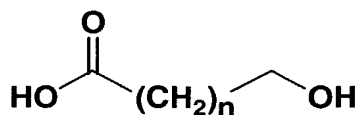wherein X comprises a biologically active molecule and Y comprises a
linker molecule that can be present or absent, under conditions suitable for
the formation of a compound having Formula 95:

95

(b) removing R6 from the compound having Formula 95 and (c) optionally
coupling a nucleotide, nucleoside, nucleic acid, oligonucleotide, amino
acid, peptide, protein, lipid, phospholipid, or label, or optionally; coupling
a compound having Formula 71 under and optionally repeating (b) and (c)
under conditions suitable for the formation of a compound having Formula
52.

166. A method for synthesizing a compound having Formula 53:

53

wherein B comprises H, a nucleoside base, or a non-nucleosidic base with
or without protecting groups; each R1 independently comprises O, N, S,
alkyl, or substituted N; each R2 independently comprises O, OH, H, alkyl,
alkylhalo, O-alkyl, O-alkylhalo, S, N, substituted N, or a phosphorus

SUBSTITUTE SHEET (RULE 26)

containing group; each R3 independently comprises N or O-N, each R4 independently comprises O, CH2, S, sulfone, or sulfoxy; X comprises H, a removable protecting group, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, or label; W comprises a linker molecule or chemical linkage that can be present or absent; SG comprises a sugar, each n is independently an integer from about 1 to about 50; and N' is an integer from about 1 to about 10, comprising: coupling a compound having Formula 73:

73

wherein R1, R2, R3, R4, X, W, B, N' and n are as defined in Formula 53, with a compound having Formula 74:
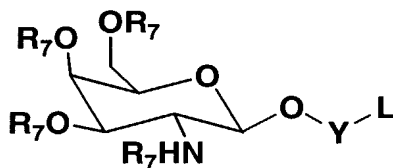
74

wherein Y comprises a linker molecule or chemical linkage that can be present or absent; L represents a reactive chemical group, and each R7 independently comprises an acyl group that can be present or absent, for example a acetyl group; under conditions suitable for the formation of a compound having Formula 53.

167.   A method for the synthesis of a compound having Formula 54:
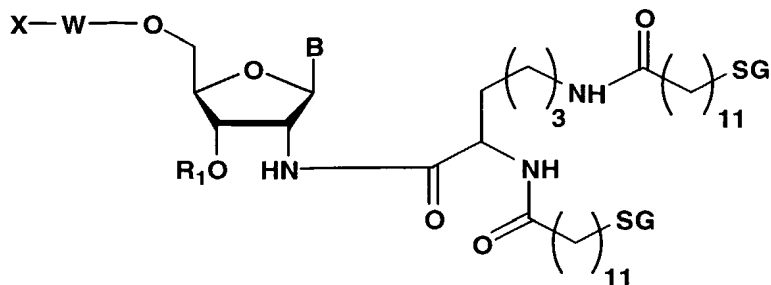
54

wherein B comprises H, a nucleoside base, or a non-nucleosidic base with
or without protecting groups; each R1 independently comprises O, OH, H,
alkyl, alkylhalo, O-alkyl, O-alkylhalo, S, N, substituted N, or a phosphorus
containing group; X comprises H, a removable protecting group, amino,
substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide,
enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid,
biologically active molecule or label; W comprises a linker molecule or
chemical linkage that can be present or absent; SG comprises a sugar,
comprising (a) coupling a compound having Formula 75:
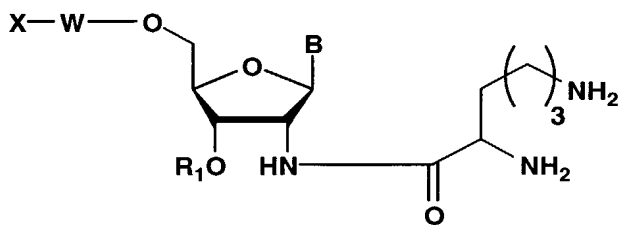
75

wherein R1, R2, R3, R4, X, W, and B are as defined in Formula 53, with a
compound having Formula 74:

74
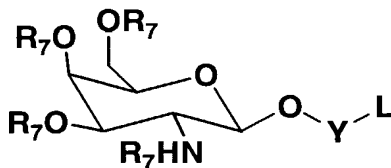
wherein Y comprises a C11 alkyl linker molecule; L represents a reactive
chemical group, for example a NHS ester, and each R7 independently

comprises an acyl group that can be present or absent, under conditions suitable for the formation of a compound having Formula 54.

168.   A method for the synthesis of a compound having Formula 55:

55

wherein each R1 independently comprises O, N, S, alkyl, or substituted N; each R2 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylhalo, S, N, substituted N, or a phosphorus containing group; each R3 independently comprises H, OH, alkyl, substituted alkyl, or halo; X comprises H, a removable protecting group, nucleotide, nucleoside, nucleic acid, oligonucleotide, or enzymatic nucleic acid or biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent; SG comprises a sugar, each n is independently an integer from about 1 to about 50; and N' is an integer from about 1 to about 100, comprising: (a) coupling a compound having Formula 76:

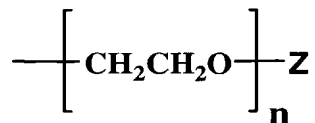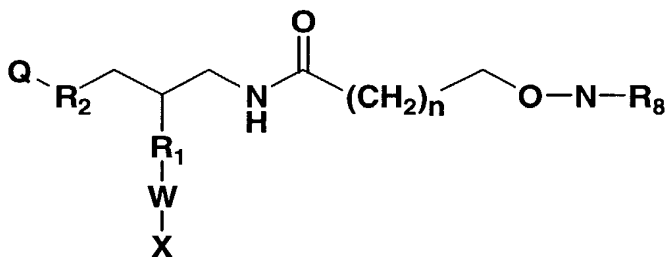

76

wherein R1 can include the groups:

and wherein R2 can include the groups:

and wherein each R3 independently comprises H, OH, alkyl, substituted alkyl, or halo; SG comprises a sugar, for example galactose, galactosamine, N-acetyl-galactosamine, glucose, mannose, fructose, or fucose and the respective D or L, alpha or beta isomers, and n is an integer from about 1 to about 20, to a compound X-W, wherein X comprises a nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, biologically active molecule or label, and W comprises a linker molecule or chemical linkage that can be present or absent; and (b) optionally repeating step (a) under conditions suitable for the formation of a compound having Formula 55.

169.   A method for the synthesis of a compound having Formula 56:

56

wherein R1 comprises H, alkyl, alkylhalo, N, substituted N, or a phosphorus containing group; R2 comprises H, O, OH, alkyl, alkylhalo, halo, S, N, substituted N, or a phosphorus containing group; X comprises H, a removable protecting group, amino, substituted amino, nucleotide,

nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, biologically active molecule or label; W comprises a linker molecule or chemical linkage that can be present or absent; SG comprises a sugar, and each n is independently an integer from about 0 to about 20, comprising: (a) coupling a compound having Formula 77:

$$X-W-O-(CH_2)_n$$
$$(CH_2)_n-NH_2$$
$$(CH_2)_n$$
$$R_2-OR_1$$

77

wherein each R1, X, W, and n are as defined in Formula 56, to a compound having Formula 74:

$$R_7O \quad OR_7$$
$$R_7O \quad O$$
$$R_7HN \quad O-Y-L$$

74

wherein Y comprises an alkyl linker molecule of length n, where n is an integer from about 1 to about 20; L represents a reactive chemical group, for example a NHS ester, and each R7 independently comprises an acyl group that can be present or absent, for example a acetyl group; and (b) optionally coupling X-W, wherein X comprises a removable protecting group, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid, amino acid, peptide, protein, lipid, phospholipid, or label and W comprises a linker molecule or chemical linkage that can be present or absent, under conditions suitable for the formation of a compound having Formula 54.

170.   A method for synthesizing a compound having Formula 57:
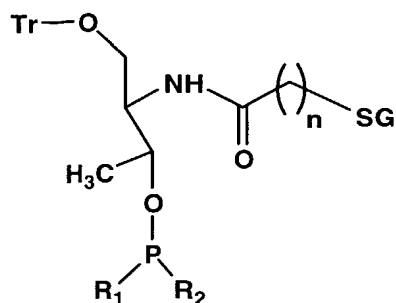
57

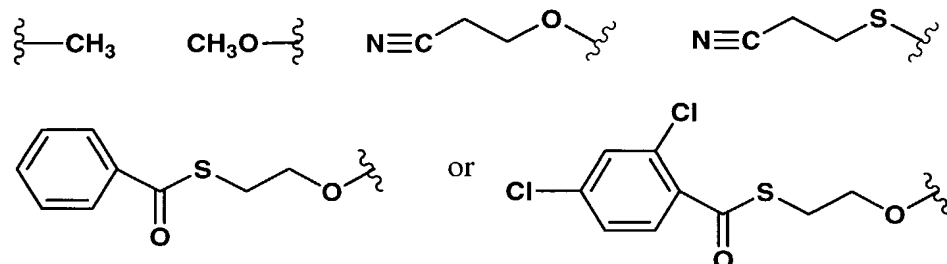wherein R1 can include the groups:

or

and wherein R2 can include the groups:

or

and wherein Tr is a removable protecting group, for example a trityl, monomethoxytrityl, or dimethoxytrityl; SG comprises a sugar, and n is an integer from about 1 to about 20, comprising: (a) coupling a compound having Formula 77:

77
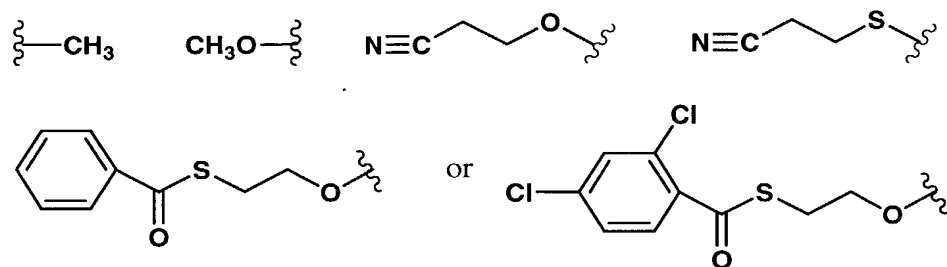
wherein R1 and X comprise H, to a compound having Formula 74:

$$R_7O \quad OR_7$$

(structure 74: sugar ring with R_7O, OR_7, R_7O, R_7HN substituents and $O\text{-}Y\text{-}L$ group)

74

wherein Y comprises an alkyl linker molecule of length n, where n is an integer from about 1 to about 20; L represents a reactive chemical group, for example a NHS ester, and each R7 independently comprises an acyl group that can be present or absent, for example a acetyl group; and (b) introducing a trityl group, for example a dimethoxytrityl, monomethoxytrityl, or trityl group to the primary hydroxyl of the product of (a) and (c) introducing a phosphorus containing group having Formula 78:

$$\begin{array}{c} R_3 \\ | \\ P \\ R_1 \quad R_2 \end{array}$$

78

wherein R1 can include the groups:

$\xi\text{—}CH_3 \qquad CH_3O\text{—}\xi \qquad N{\equiv}C\text{—}\wedge\text{—}O\text{—}\sim \qquad N{\equiv}C\text{—}\wedge\text{—}S\text{—}\sim$

(benzoyl-S-CH2CH2-O structure)   or   (2,4-dichlorobenzoyl-S-CH2CH2-O structure)

and wherein each R2 and R3 independently can include the groups:

$\xi\text{—N}\big(CH_3\big)_2 \qquad \xi\text{—N}\big(CH_2CH_3\big)\big(CH_2CH_3\big) \qquad \xi\text{—N}\big(isopropyl\big)_2 \qquad \xi\text{—N(pyrrolidine)} \qquad$ or $\qquad \xi\text{—N(morpholine)}$

to the secondary hydroxyl of the product of (b) under conditions suitable for the formation of a compound having Formula 57.

171.   In another embodiment, the invention features a method for synthesizing a
        compound having Formula 60:
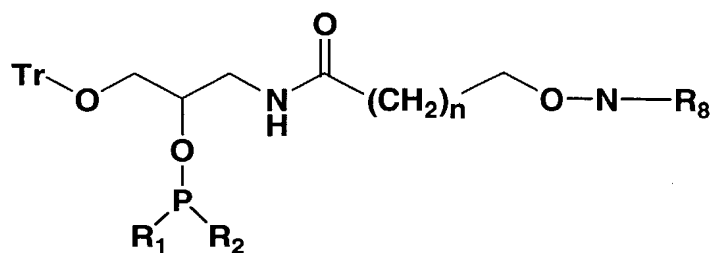
60

5   wherein R1 can include the groups:
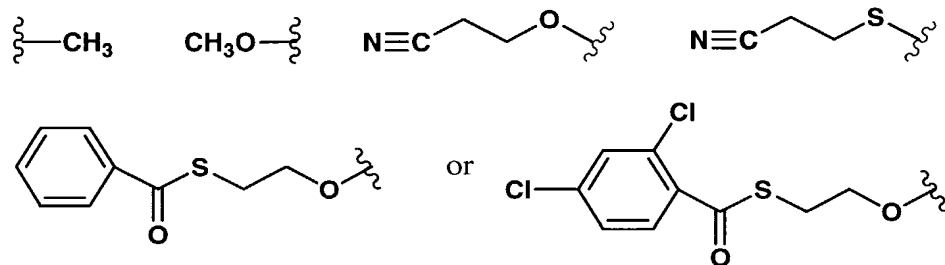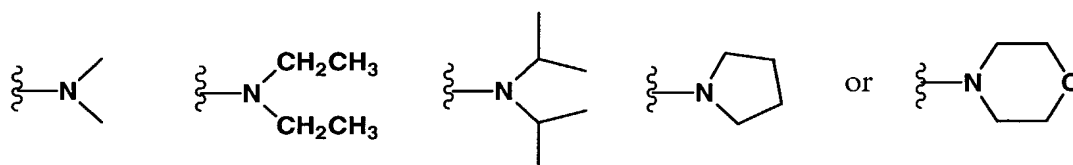
        and wherein R2 can include the groups:

        and wherein Tr is a removable protecting group, for example a trityl,
10   monomethoxytrityl, or dimethoxytrityl; n is an integer from about 1 to
        about 50; and R8 is a nitrogen protecting group, comprising: (a)
        introducing carboxy protection to a compound having Formula 79:

79

15   wherein n is an integer from about 1 to about 50, under conditions suitable
        for the formation of a compound having Formula 80:

180

$$R_7O \overset{\overset{\displaystyle O}{\|}}{} (CH_2)_n \diagdown OH$$

80

wherein n is an integer from about 1 to about 50 and R7 is a carboxylic acid protecting group, for example a benzyl group; (b) introducing a nitrogen containing group to the product of (a) under conditions suitable for the formation of a compound having Formula 81:

$$R_7 \overset{\overset{\displaystyle O}{\|}}{} (CH_2)_n \diagdown O-N-R_8$$

81

wherein n and R7 are as defined in Formula 80 and R8 is a nitrogen protecting group, for example a phthaloyl, trifluoroacetyl, FMOC, or monomethoxytrityl group; (c) removing the carboxylic acid protecting group from the product of (b) and introducing aminopropanediol under conditions suitable for the formation of a compound having Formula 82:

$$HO \diagdown \underset{\underset{\displaystyle HO}{|}}{} \diagdown NH \overset{\overset{\displaystyle O}{\|}}{} (CH_2)_n \diagdown O-N-R_8$$

82

wherein n and R8 are as defined in Formula 81; (d) introducing a removable protecting group to the product of (c) under conditions suitable for the formation of a compound having Formula 83:

$$TrO \diagdown \underset{\underset{\displaystyle HO}{|}}{} \diagdown NH \overset{\overset{\displaystyle O}{\|}}{} (CH_2)_n \diagdown O-N-R_8$$

83

wherein Tr, n and R8 are as defined in Formula 60; and (e) introducing a phosphorus containing group having Formula 78:

SUBSTITUTE SHEET (RULE 26)

78
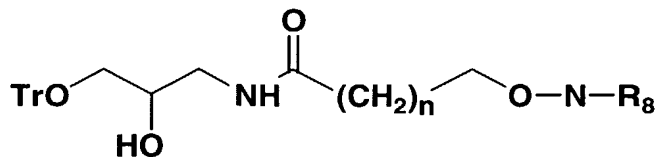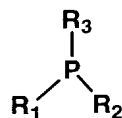
wherein R1 can include the groups:

and wherein each R2 and R3 independently can include the groups:

to the product of (d) under conditions suitable for the formation of a compound having Formula 60.

172.   A method for the synthesis of a compound having Formula 59:

59

wherein each R1 independently comprises O, S, N, substituted N, or a phosphorus containing group; each R2 independently comprises O, S, or N; X comprises H, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, enzymatic nucleic acid or biologically active molecule; n is an integer from about 1 to about 50, Q comprises H or a removable protecting group which can be optionally absent, each W

independently comprises a linker molecule or chemical linkage that can be present or absent, and V comprises a protein or peptide or a compound having Formula 45:

$$-\left[CH_2CH_2O\right]_n-Z$$

45

wherein Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; and n is an integer from about 1 to about 100, comprising: (a) removing R8 from a compound having Formula 84:

84

wherein Q, X, W, R1, R2, and n are as defined in Formula 59 and R8 is a nitrogen protecting group under conditions suitable for the formation of a compound having Formula 85:

85

wherein Q, X, W, R1, R2, and n are as defined in Formula 59; (b) introducing a group V to the product of (a) via the formation of an oxime

linkage, wherein V comprises a protein or peptide or a compound having Formula 45:

$$-\left[-CH_2CH_2O-\right]_n Z$$
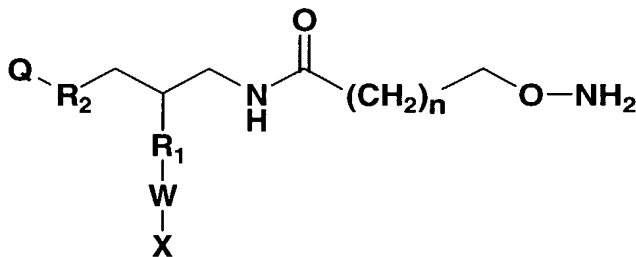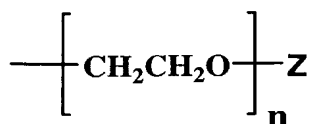
45

5   wherein Z comprises H, OH, O-alkyl, SH, S-alkyl, alkyl, substituted alkyl, aryl, substituted aryl, amino, substituted amino, nucleotide, nucleoside, nucleic acid, oligonucleotide, amino acid, peptide, protein, lipid, phospholipid, or label; and n is an integer from about 1 to about 100, under conditions suitable for the formation of a compound having Formula 59.

10   173.   A method for synthesizing a compound having Formula 64:

64

wherein X comprises a biologically active molecule; each W independently comprises a linker molecule or chemical linkage that can be present or absent, Y comprises a linker molecule that can be present or absent; each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N, A comprises a nitrogen containing group, and B comprises a lipophilic group, comprising: (a) introducing a compound having Formula 66:

66

wherein R1 is defined as in Formula 64 and can include the groups:

ξ—CH₃   CH₃O—ξ N≡C⌒⌒O⌒ξ N≡C⌒⌒S⌒ξ

[benzoyl-S-ethyl-O structure]   or   [dichlorobenzoyl-S-ethyl-O structure]

and wherein R2 is defined as in Formula 64 and can include the groups:

ξ—N(CH₃)₂ ξ—N(CH₂CH₃)(CH₂CH₃) ξ—N(iPr)₂ ξ—N(pyrrolidine)   or ξ—N(morpholine)

5   and wherein each R5 independently comprises O, N, or S and each R6 independently comprises a removable protecting group to a compound having Formula 67:

X——W——Y

67

10   wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent, and Y comprises a linker molecule that can be present or absent, under conditions suitable for the formation of a compound having Formula 68:

$$X——W——Y—R_1-\overset{\overset{R_4}{\|}}{\underset{\underset{R_2}{|}}{P}}-R_3 \begin{cases} R_5——R_6 \\ \\ R_5——R_6 \end{cases}$$

15   68

wherein X comprises a biologically active molecule; W comprises a linker molecule or chemical linkage that can be present or absent, Y comprises a linker molecule that can be present or absent; and each R1, R2, R3, and R4 independently comprises O, OH, H, alkyl, alkylhalo, O-alkyl, O-

20   alkylcyano, S, S-alkyl, S-alkylcyano, N or substituted N comprising, each

R5 independently comprises O, S, or N; and each R6 is independently a removable protecting group; (b) removing R6 from the compound having Formula 26 and (c) introducing a compound having Formula 69:

$$R_1-\underset{\underset{R_2}{|}}{P}-R_3-W-B$$

5

69

wherein R1 is defined as in Formula 64 and can include the groups:

and wherein R2 is defined as in Formula 64 and can include the groups:

10   and wherein R3, W and B are defined as in Formula 64; and introducing a compound having Formula 69':
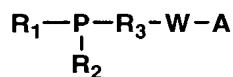
$$R_1-\underset{\underset{R_2}{|}}{P}-R_3-W-A$$

69'

wherein R1 is defined as in Formula 64 and can include the groups:

15

and wherein R2 is defined as in Formula 48 and can include the groups:

SUBSTITUTE SHEET (RULE 26)

$$\text{Y—W—}\underset{\underset{\text{R}_1}{|}}{\text{C}}\text{=N—O—X}$$
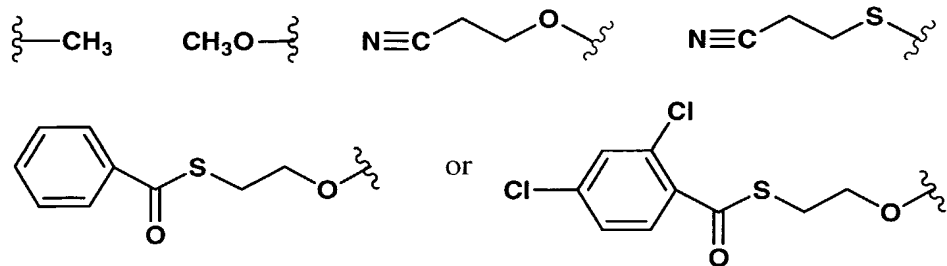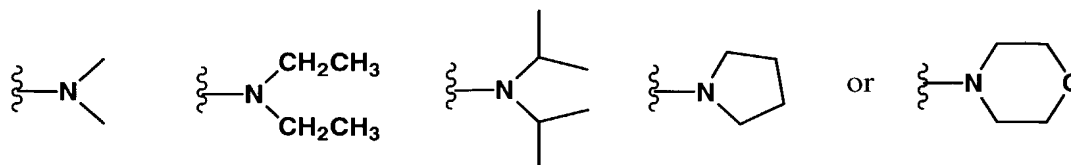
and wherein R3, W and A are defined as in Formula 64; to the compound having Formula 68 under conditions suitable for the formation of a compound having Formula 64.

5 174. In another embodiment, the invention features a method for the synthesis of a compound having Formula 87:

$$\text{Y——W——}\underset{\underset{\text{R}_1}{|}}{\text{C}}\text{==N—O—X}$$
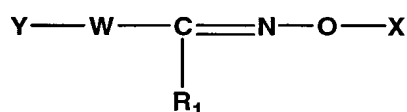
87

10 wherein X comprises a protein, peptide, antibody, lipid, phospholipid, oligosaccharide, label, biologically active molecule, W comprises a linker molecule or chemical linkage that can be present or absent, Y comprises a biologically active molecule, and R1 comprises H, alkyl, or substituted alkyl, comprising (a) coupling a compound having Formula 89:

$$\text{Y——W——}\underset{\underset{\text{R}_1}{|}}{\text{C}}\text{==O}$$

15 89

wherein Y, W and R are as defined in Formula 87, with a compound having Formula 90:

$$\text{H}_2\text{N—O—X}$$
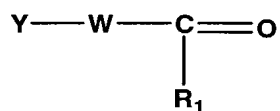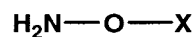
90

20 wherein X is as defined in Formula 87, under conditions suitable for the formation of a compound having Formula 87, for example by post-synthetic conjugation of a compound having Formula 89 with a compound having Formula 90, wherein X of compound 90 comprises an enzymatic nucleic acid molecule and Y of Formula 89 comprises a peptide.

175. In another embodiment, the invention features a method for the synthesis of a compound having Formula 88:

$$Y—W—\overset{\overset{\displaystyle O}{\|}}{C}—NH—O—X$$

88
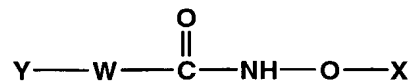
wherein X comprises a protein, peptide, antibody, lipid, phospholipid, oligosaccharide, label, biologically active molecule, W comprises a linker molecule or chemical linkage that can be present or absent, and Y comprises a biologically active molecule, comprising (a) coupling a compound having Formula 91:

$$Y—W—\underset{\underset{\displaystyle H}{|}}{C}{=}O$$
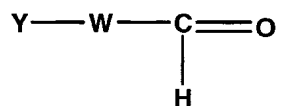
91

wherein Y and W is as defined in Formula 88, with a compound having Formula 90:
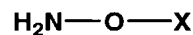
$$H_2N—O—X$$

90

wherein X is as defined in Formula 88, under conditions suitable for the formation of a compound having Formula 88, by post-synthetic conjugation of a compound having Formula 91 with a compound having Formula 90, wherein X of compound 90 comprises an enzymatic nucleic acid molecule and Y of Formula 91 comprises a peptide.

176. In one embodiment, the invention features a compound having Formula 94:

$$X—Y—W—Y—Z$$

94

wherein X comprises a protein, peptide, antibody, lipid, phospholipid, oligosaccharide, label, biologically active molecule, for example a vitamin such as folate, vitamin A, E, B6, B12, coenzyme, antibiotic, antiviral, nucleic acid, nucleotide, nucleoside, or oligonucleotide such as an

enzymatic nucleic acid, allozyme, antisense nucleic acid, siRNA, 2,5-A chimera, decoy, aptamer or triplex forming oligonucleotide, or polymers such as polyethylene glycol; each Y independently comprises a linker or chemical linkage that can be present or absent, W comprises a biodegradable nucleic acid linker molecule, and Z comprises a biologically active molecule, for example an enzymatic nucleic acid, allozyme, antisense nucleic acid, siRNA, 2,5-A chimera, decoy, aptamer or triplex forming oligonucleotide, peptide, protein, or antibody.

177. The kit of claim 97, wherein said sample is from a cancer cell.

178. The kit of claim 97, wherein said sample is from a virus infected cell.

Figure 1: Examples of Nuclease Stable Ribozyme Motifs
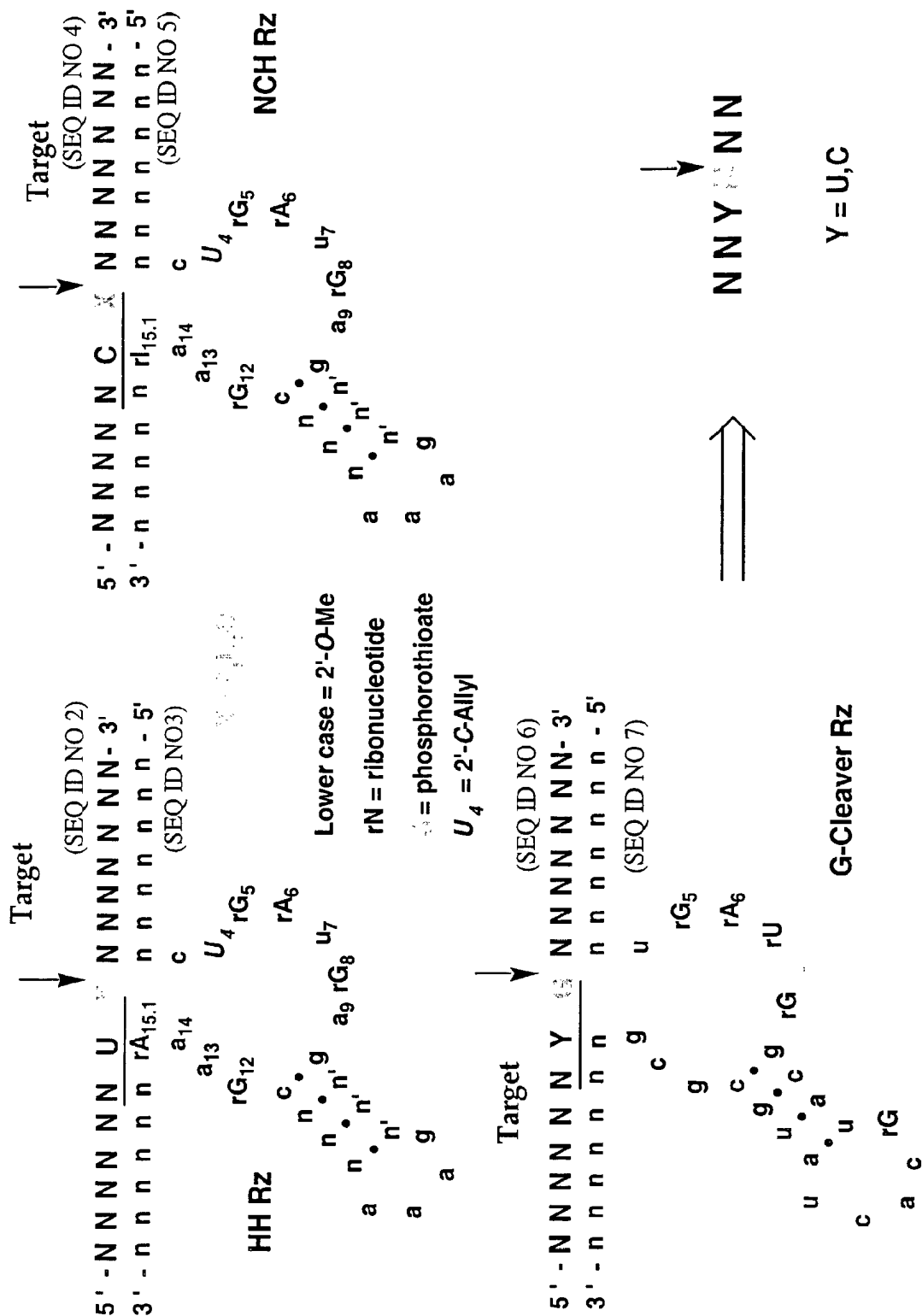
Figure 2: 2'-O-Me substituted Amberzyme Enzymatic Nucleic Acid Motif

U,C = 2'-NH₂-U,C

Lower case = 2'-O-Me

Uppercase = Ribo
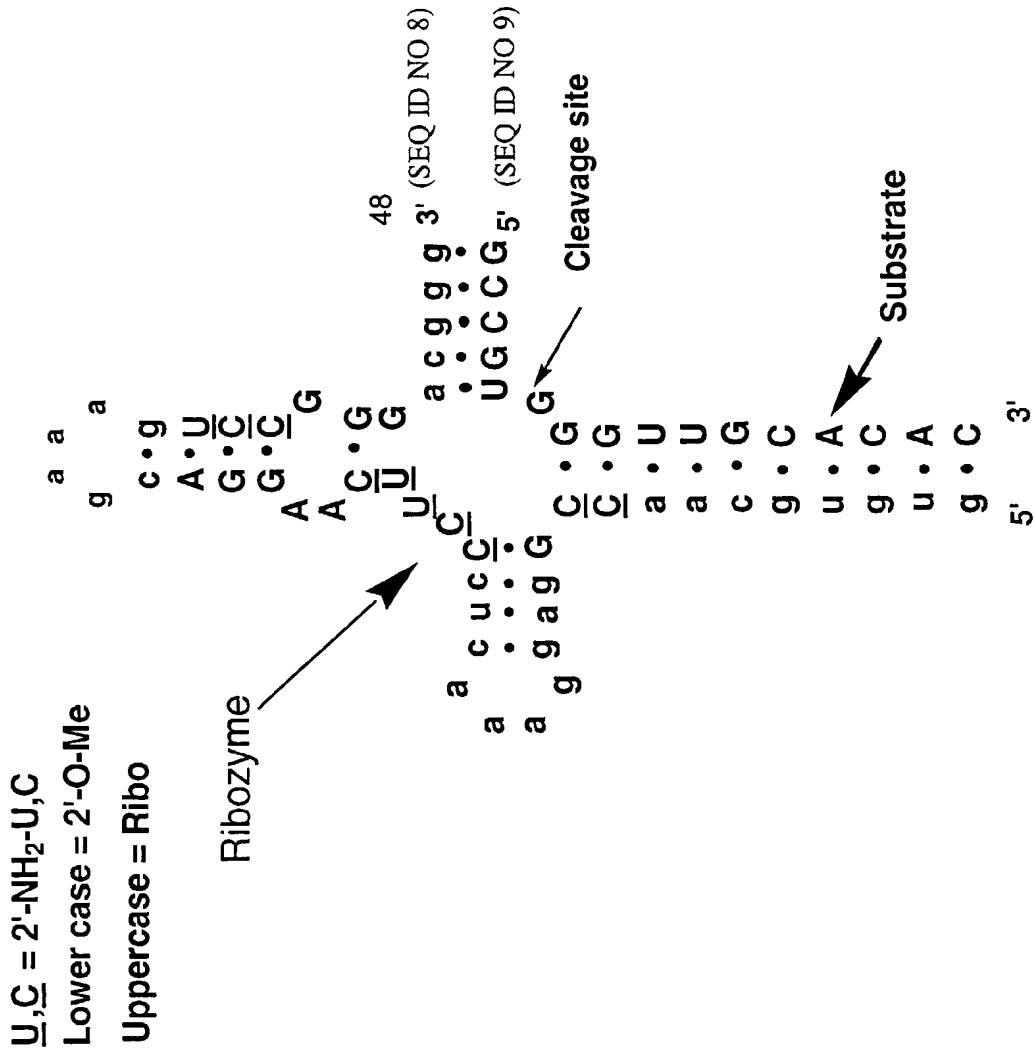
Figure 3: Stabilized Zinzyme Ribozyme Motif

Zinzyme

```
                          u  G   a
                             G   a
                         a a  a u  G
                        a    g G  C   u    (SEQ ID NO 10)
                             g    C  G
                        5'  g c  c  C  G
                     gu g c u g   g c c g c g c - B  3'
                 5'   s s s s
                 3'   CAC GGAC  G  CGGC GCG   5'
                                  ↗            (SEQ ID NO 11)
                                 Target
```

Legend

Uppercase indicates natural ribo residues

<u>C</u> indicates 2'- dexoy-2'-amino Cytidine

Lowercase indicates 2'-O-methyl

S indicates phosphorothioate/phosphorodithioate

B: 3'-3' abasic moiety
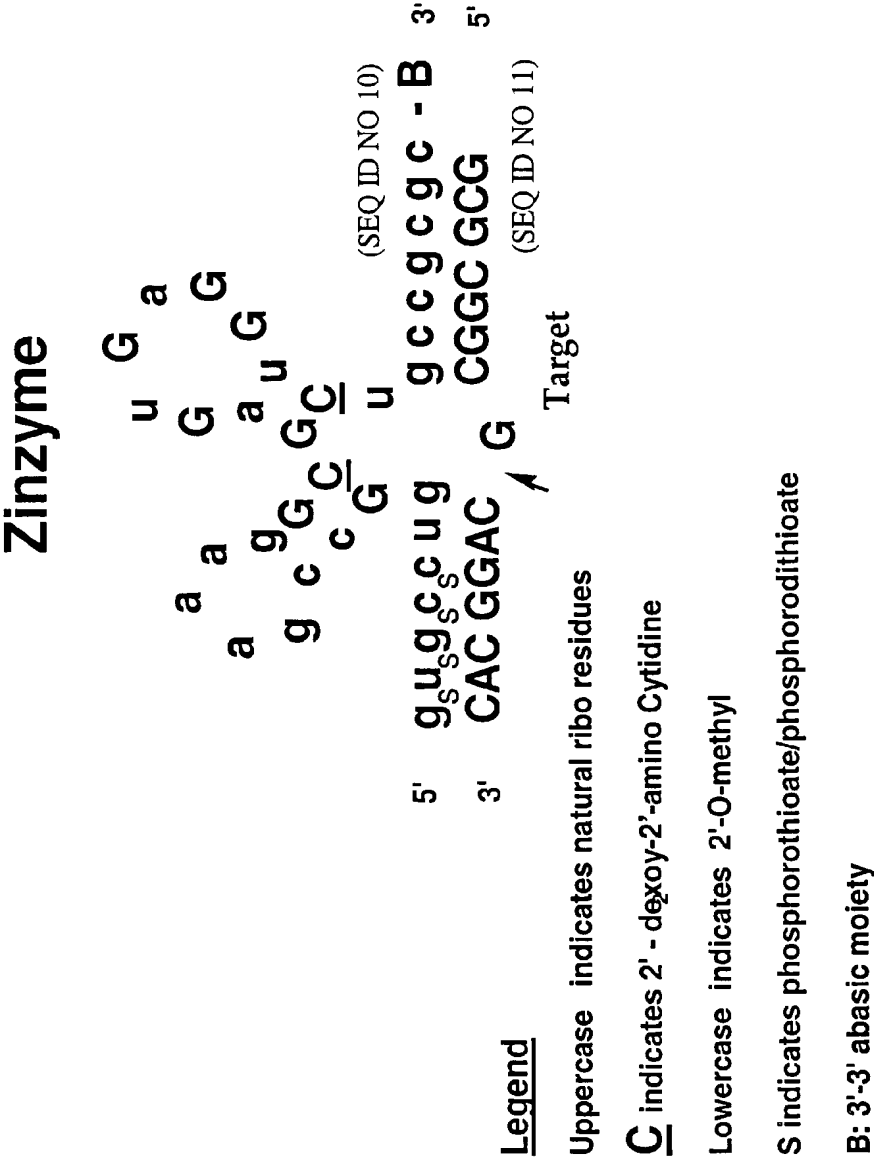
Figure 4: DNAzyme Motif
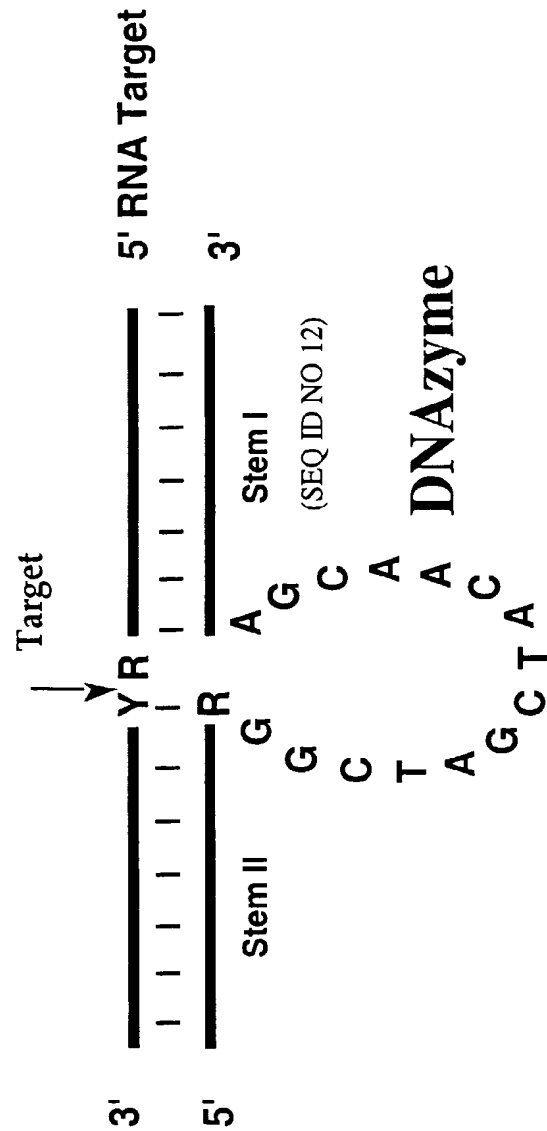
Figure 5: Synthesis of Folate Linked phosphoramidite
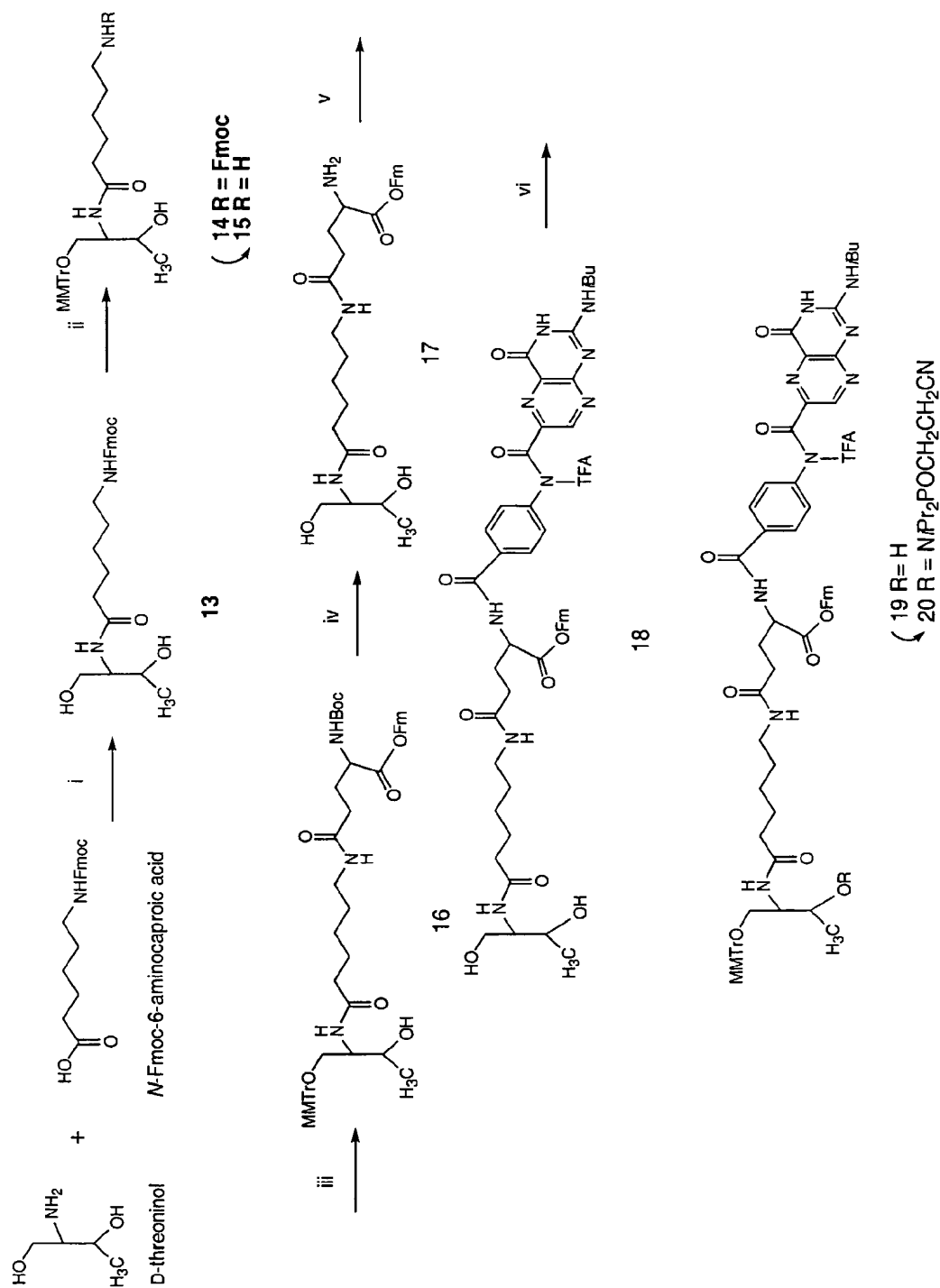
Figure 6: Fludarabine-Folate conjugates
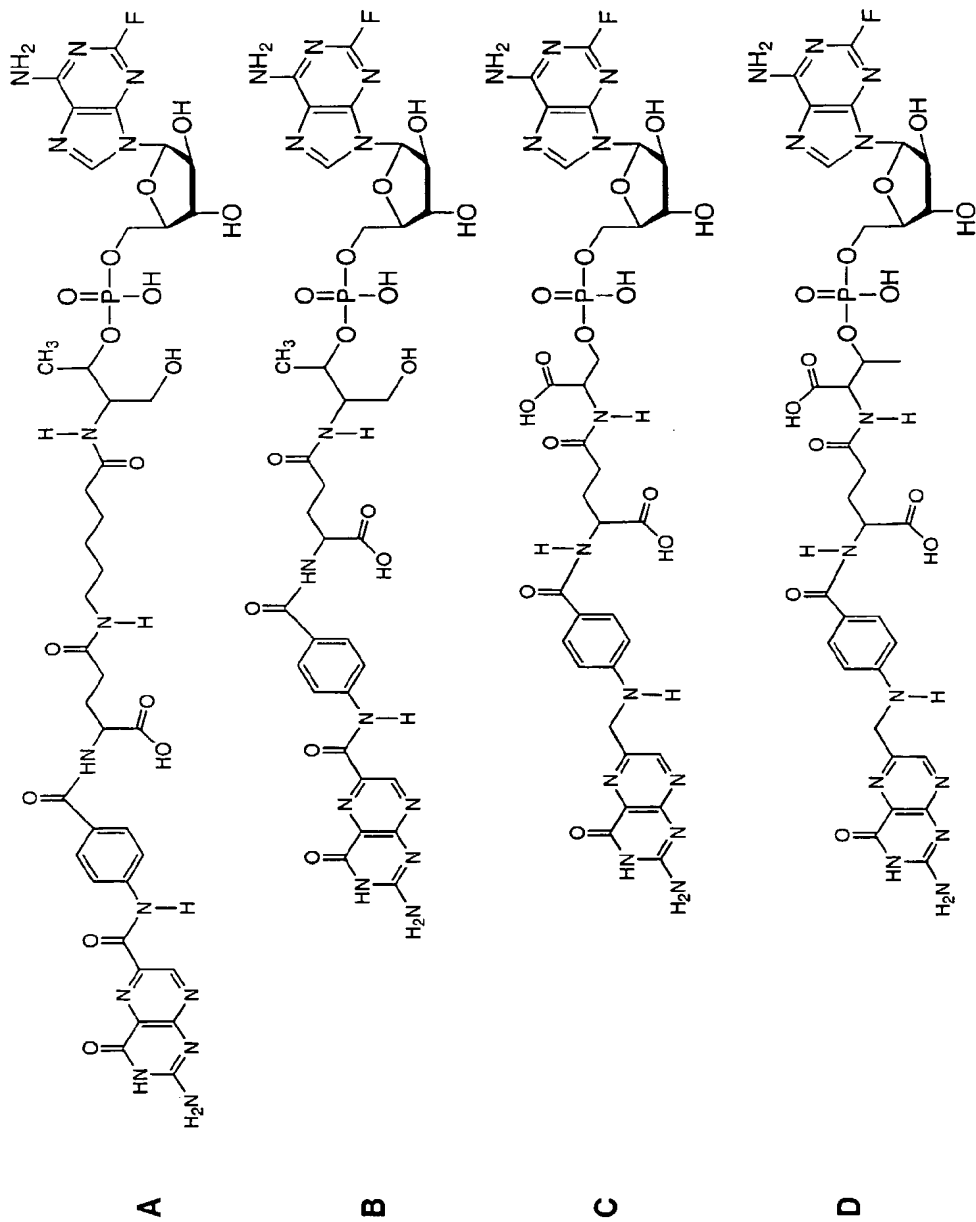
Figure 7: Solid Phase Post-synthetic conjugation of pteroic acid
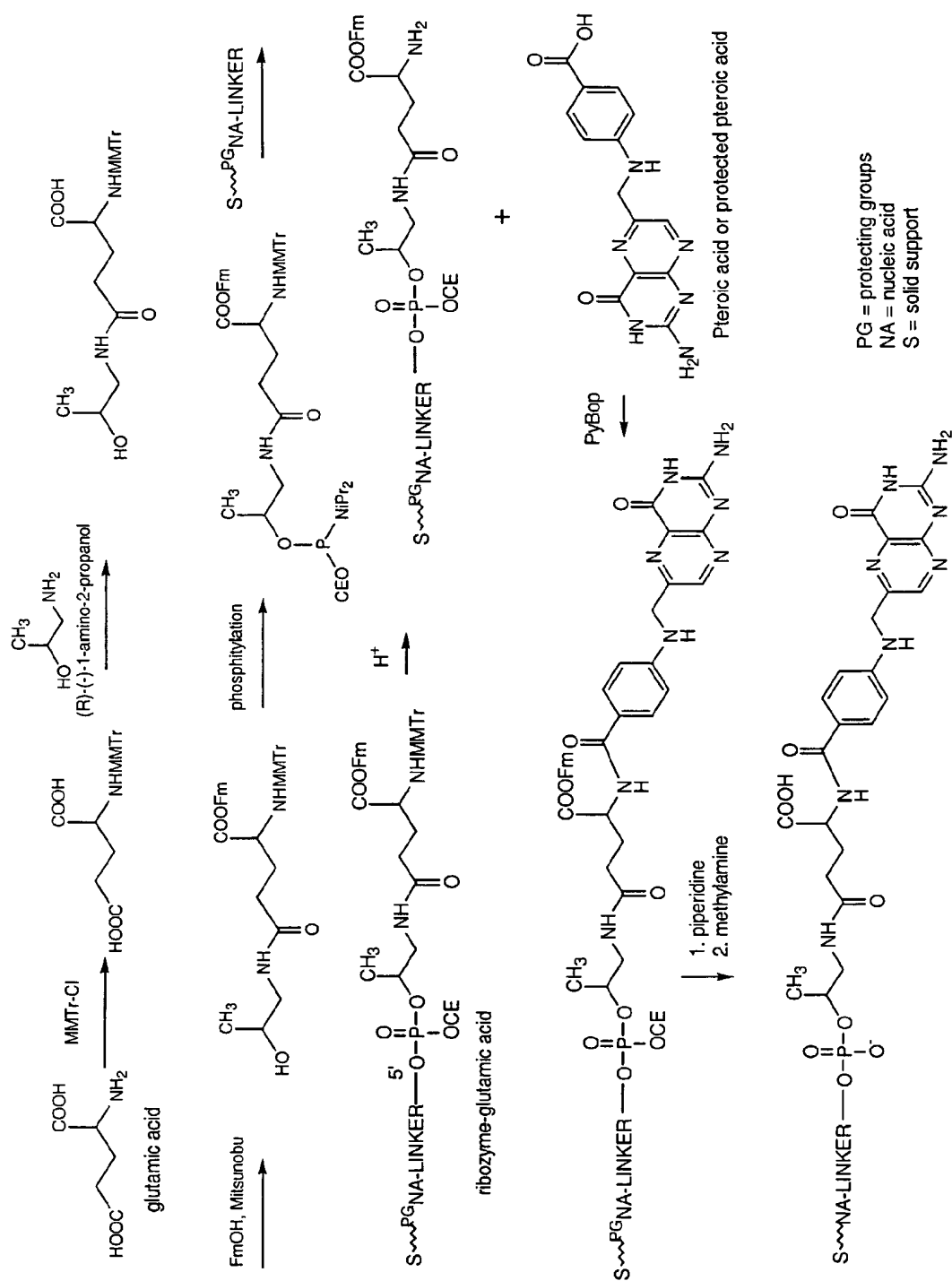
Figure 8: Chemo-enzymatic synthesis of pteroic acid synthon
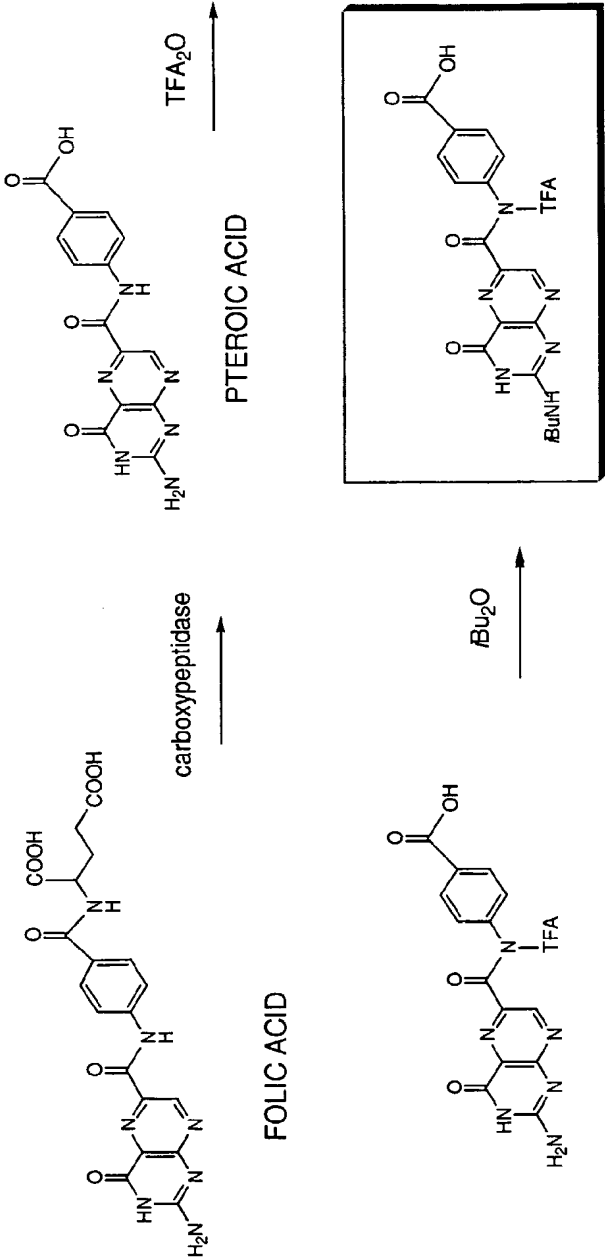
Figure 9
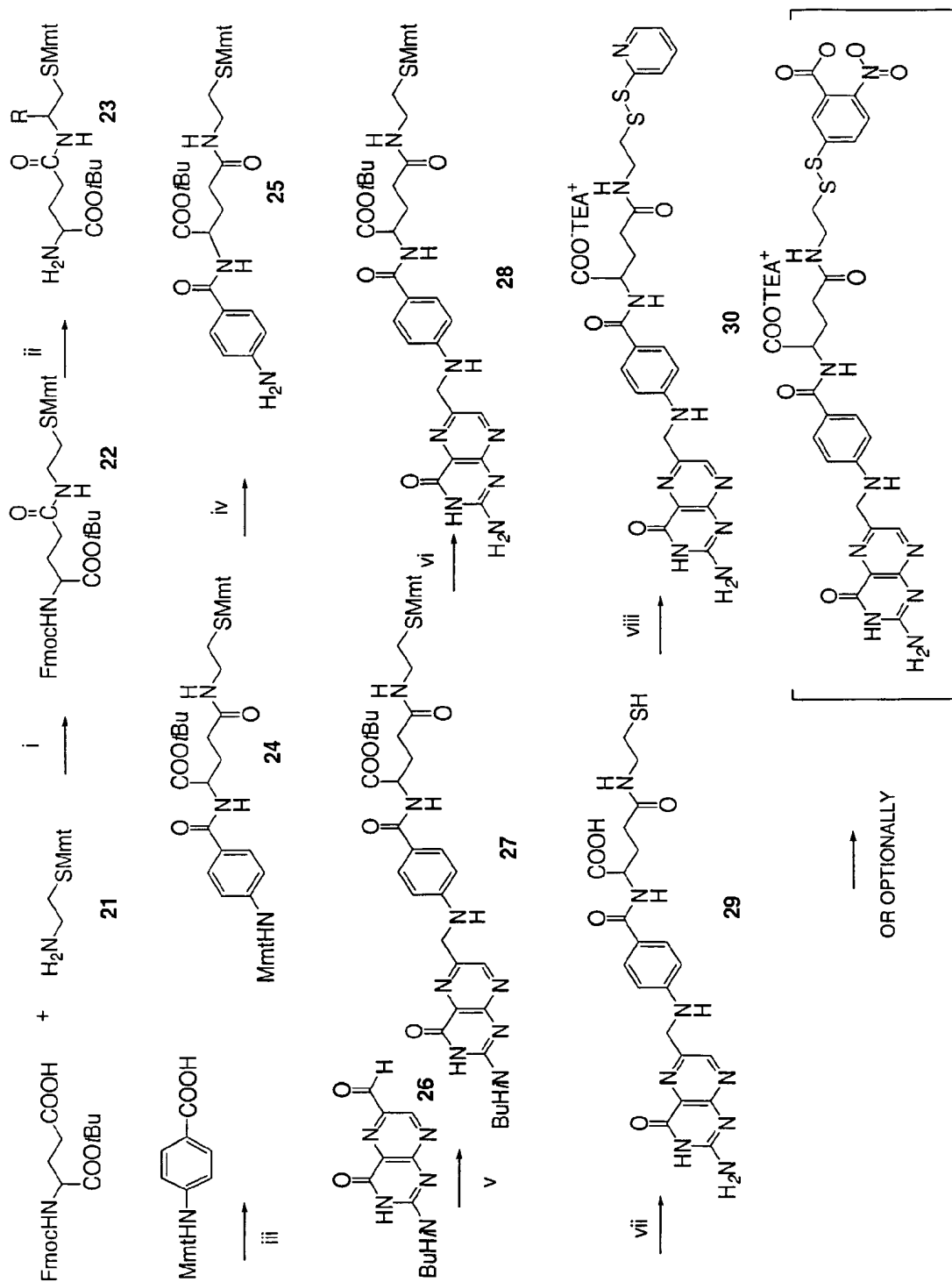
Figure 10
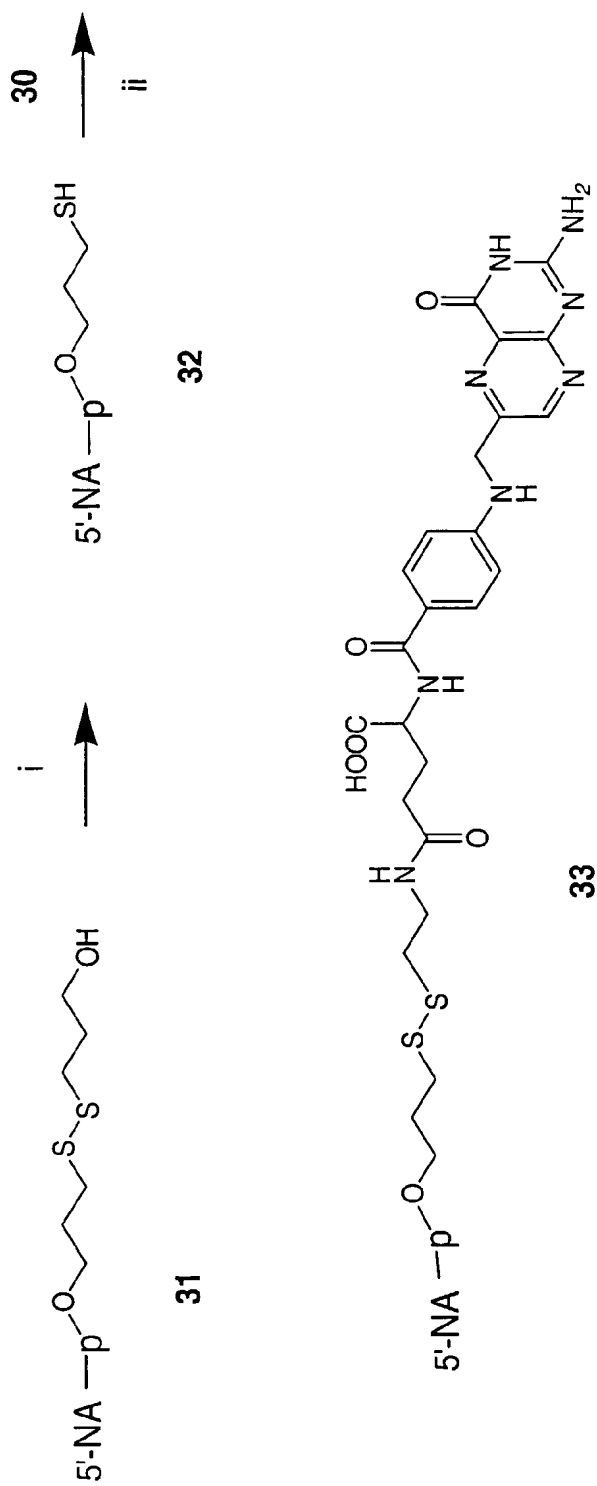
NA = Nucleic Acid, such as enzymatic nucleic acid
p = phosphorous moiety

Figure 11
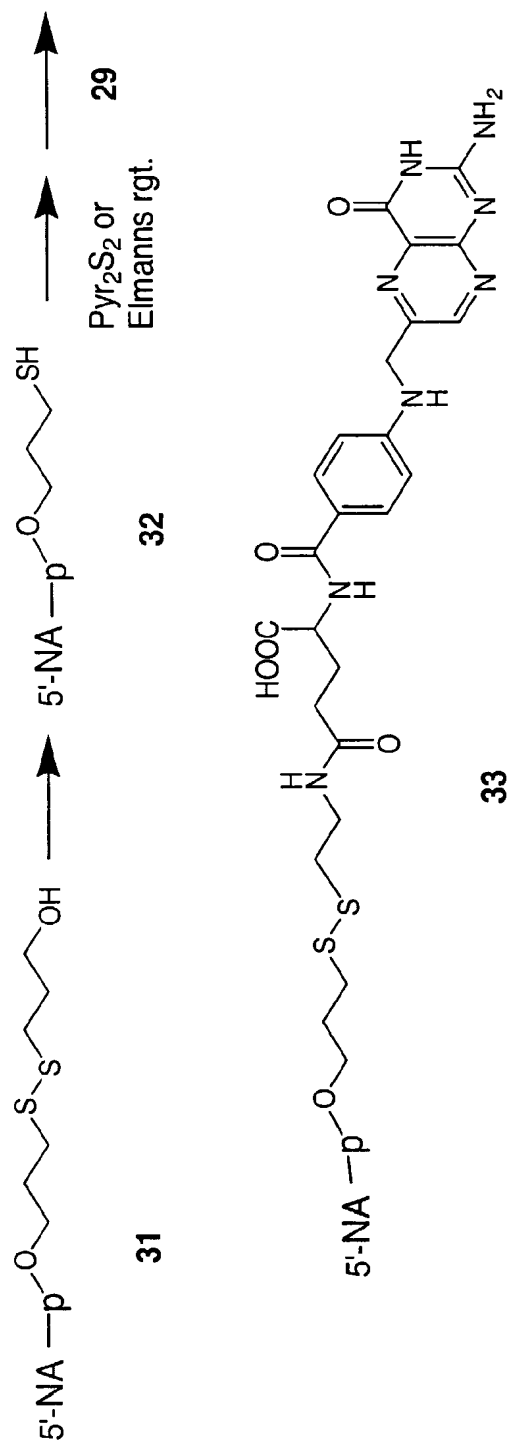
NA = Nucleic Acid, such as enzymatic nucleic acid
p = phosphorous moiety

Figure 12: Solid Phase Post-synthetic conjugation of pteroic acid

R = Ac, *i*Bu
PG = protecting groups
NA = nucleic acid
S = solid support
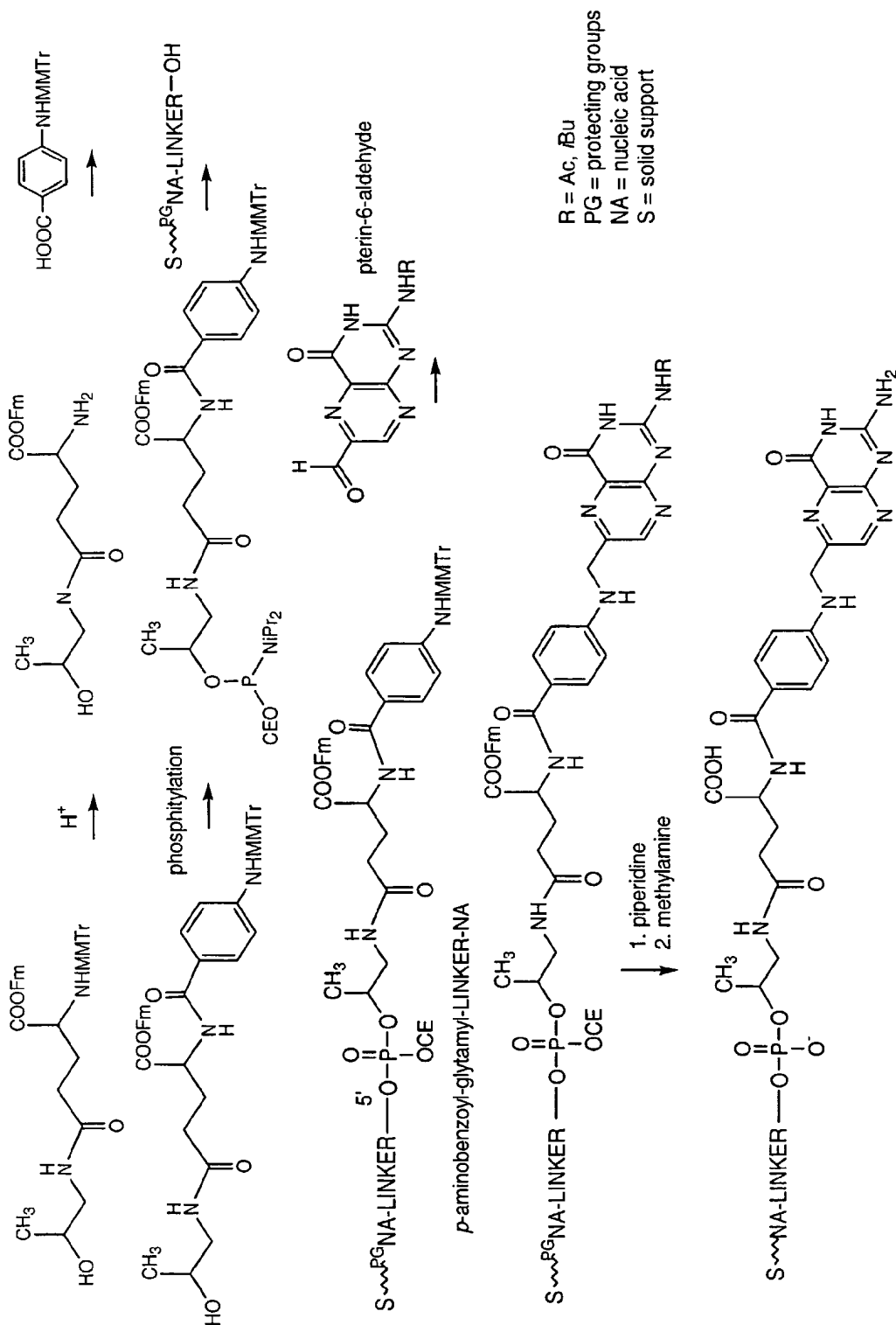
Figure 13: Synthesis of N-acetyl-D-galactosamine-2'-aminouridine conjugate
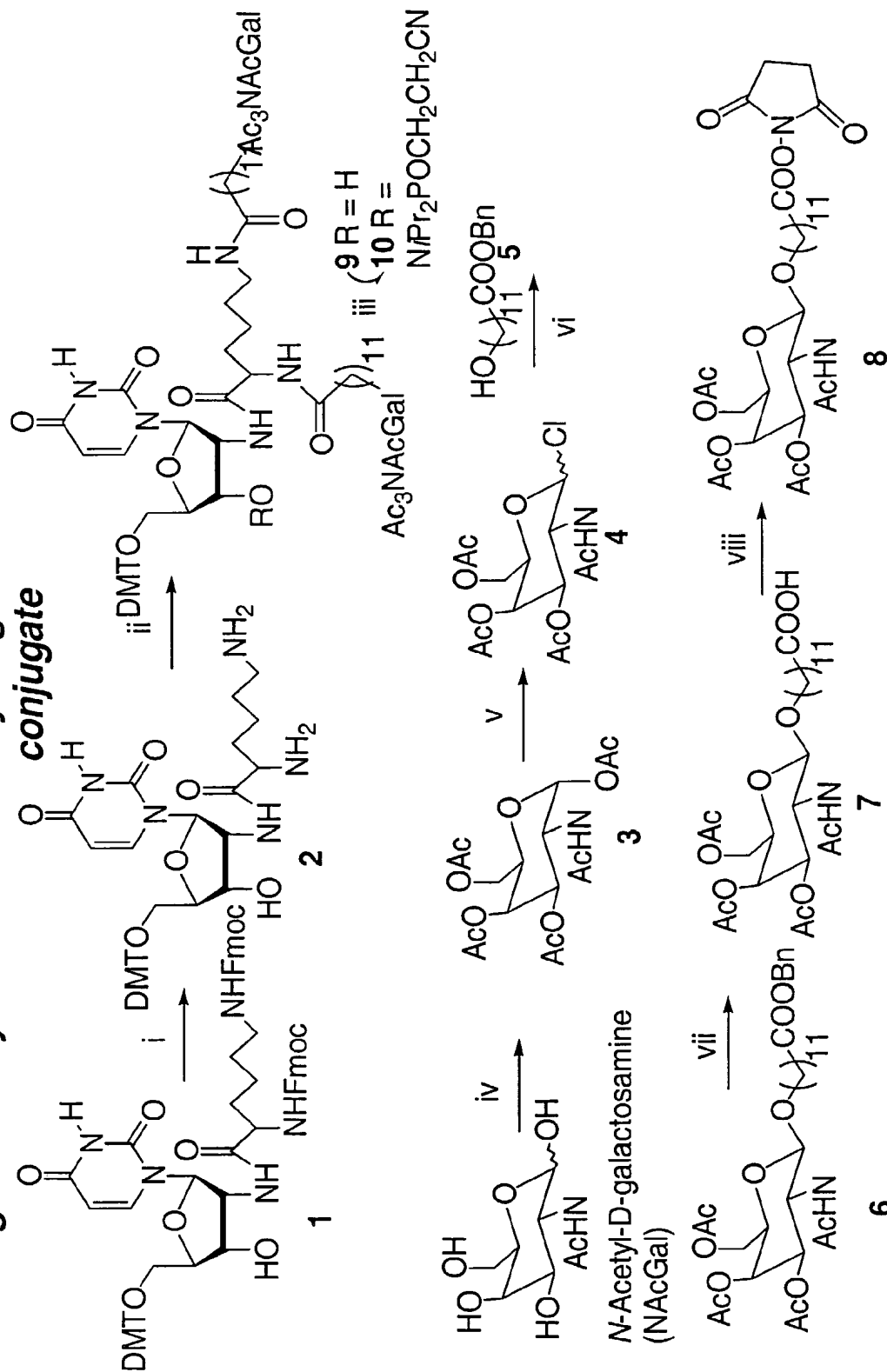
Reagents and Conditions: (i) diethylamine, DMF, (ii) **8**, diisopropylethylamine, DMF, (iii) 2-cyanoethyl N,N-diisopropylchlorophosphoramidite, 1-methylimidazole, DIPEA, CH$_2$Cl$_2$, (iv) Ac$_2$O, TEA, CH$_3$CN, (v) HCl, Ac$_2$O, (vi) Hg(CN)$_2$, MS 4A, CH$_3$NO$_2$-toluene 1:1, (vii) H$_2$, 5% Pd-C, ethanol, (viii) N-hydroxysuccinimide, DCC, THF.

Figure 14: Synthesis of N-acetyl-D-galactosamine-D-threoninol conjugate
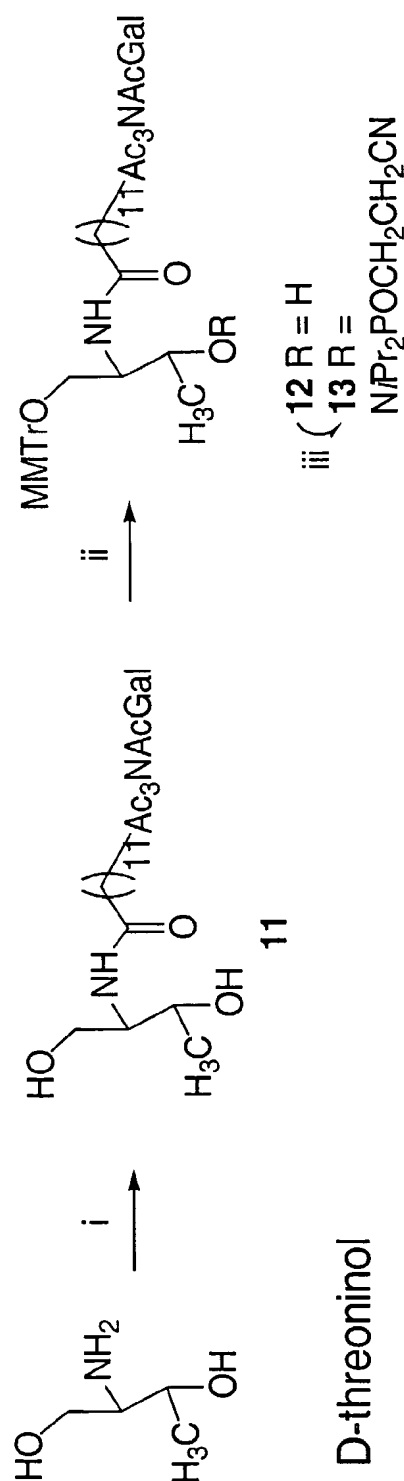
Reagents and Conditions: (i) **7**, DCC, *N*-hydroxysuccinimide, (ii) MMTr-Cl, pyridine, (iii) 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite, 1-methylimidazole, DIPEA, CH_2Cl_2.

Figure 15: Conjugation of targeting ligands to the 5'-end of a Ribozyme

N-acetyl-D-galactosamine conjugate
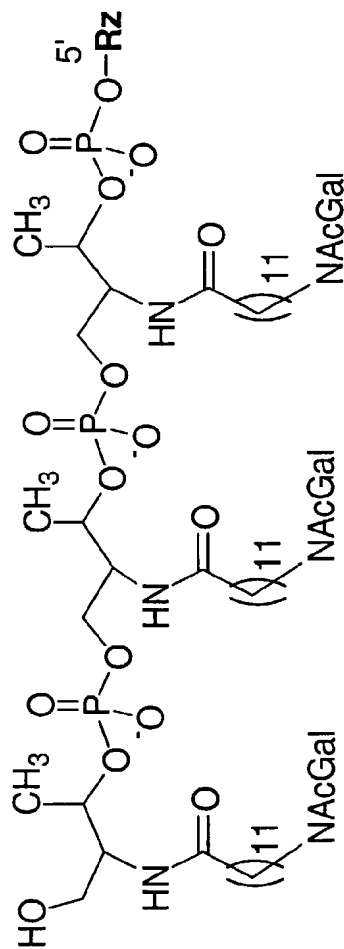
Figure 16: Synthesis of dodecanoic acid linker
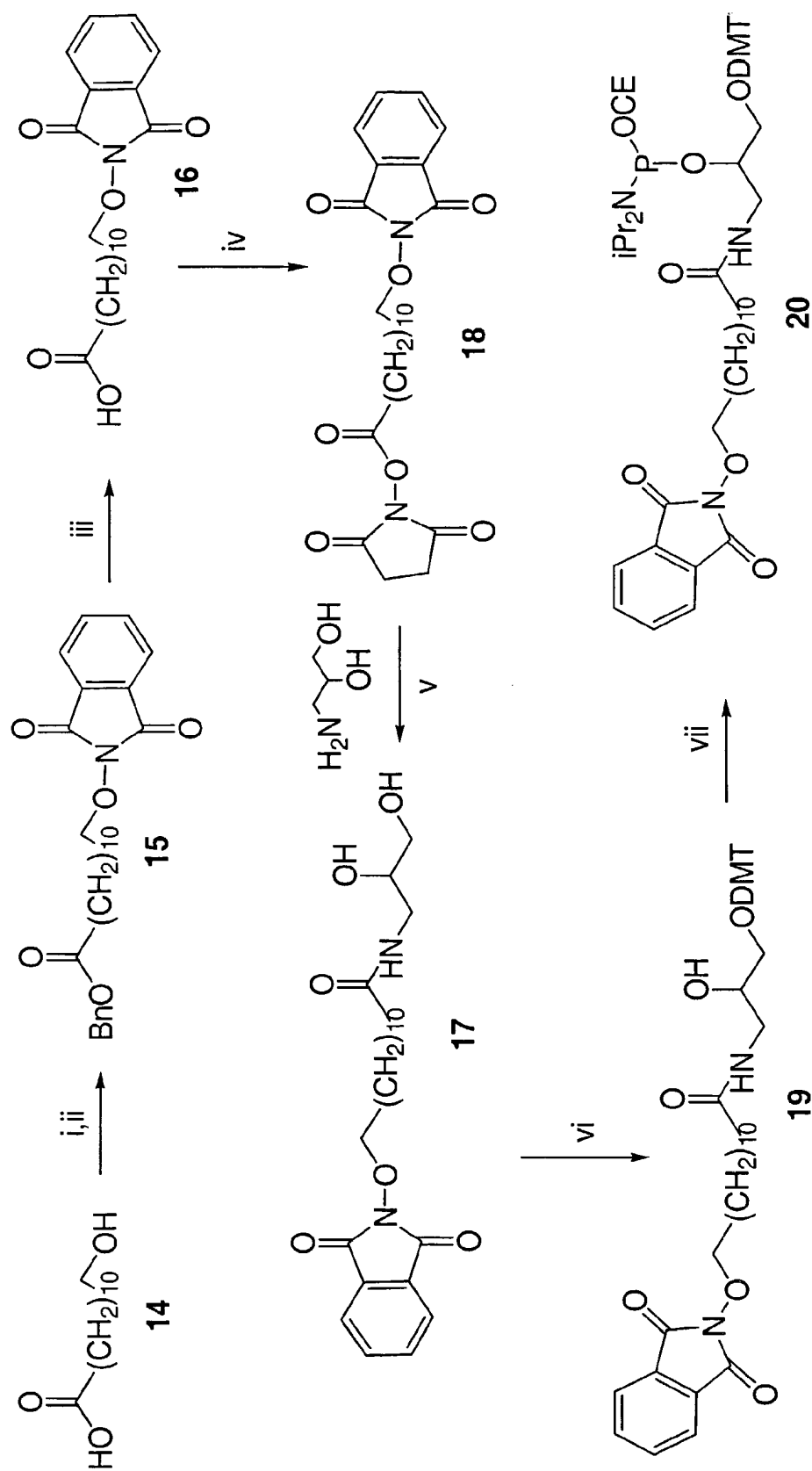
Figure 17: Oxime linked Nucleic Acid/Peptide Conjugate
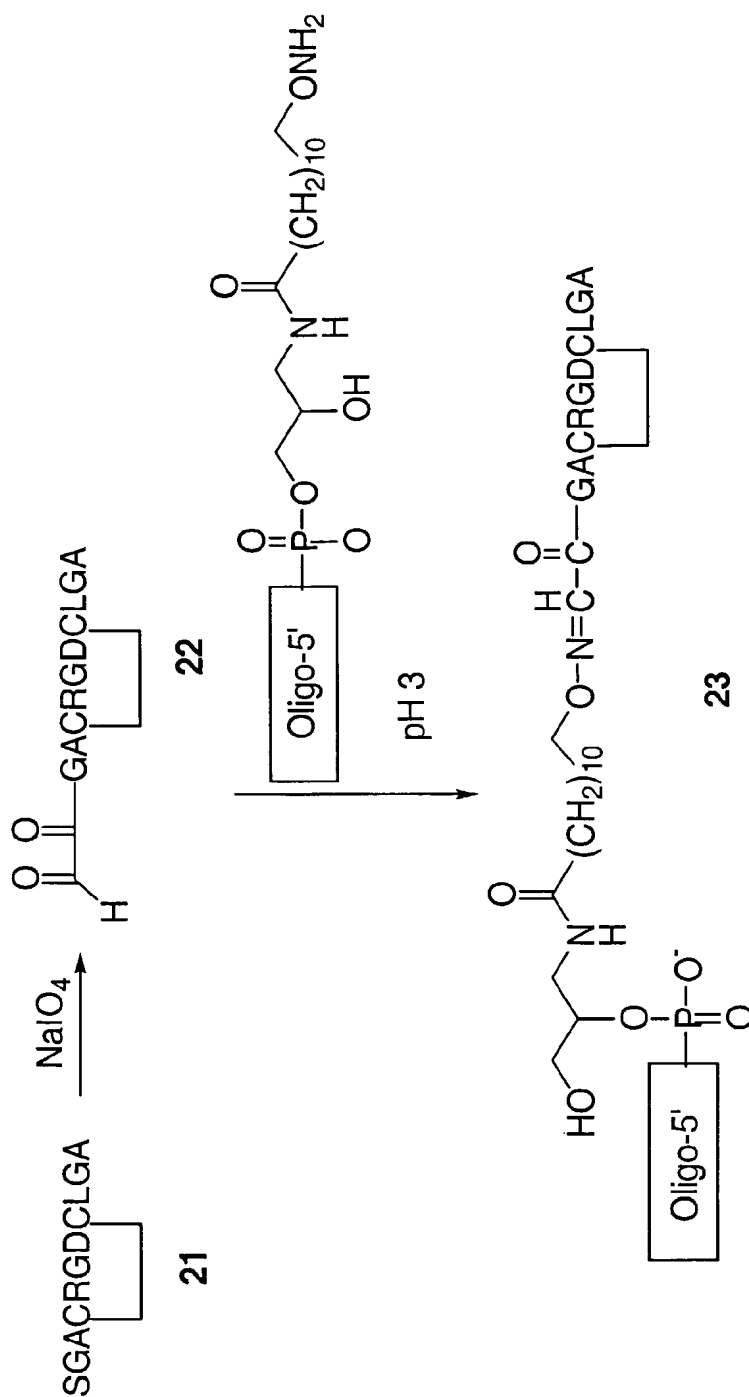
Figure 18: Nucleic Acid/Phospholipid Conjugates
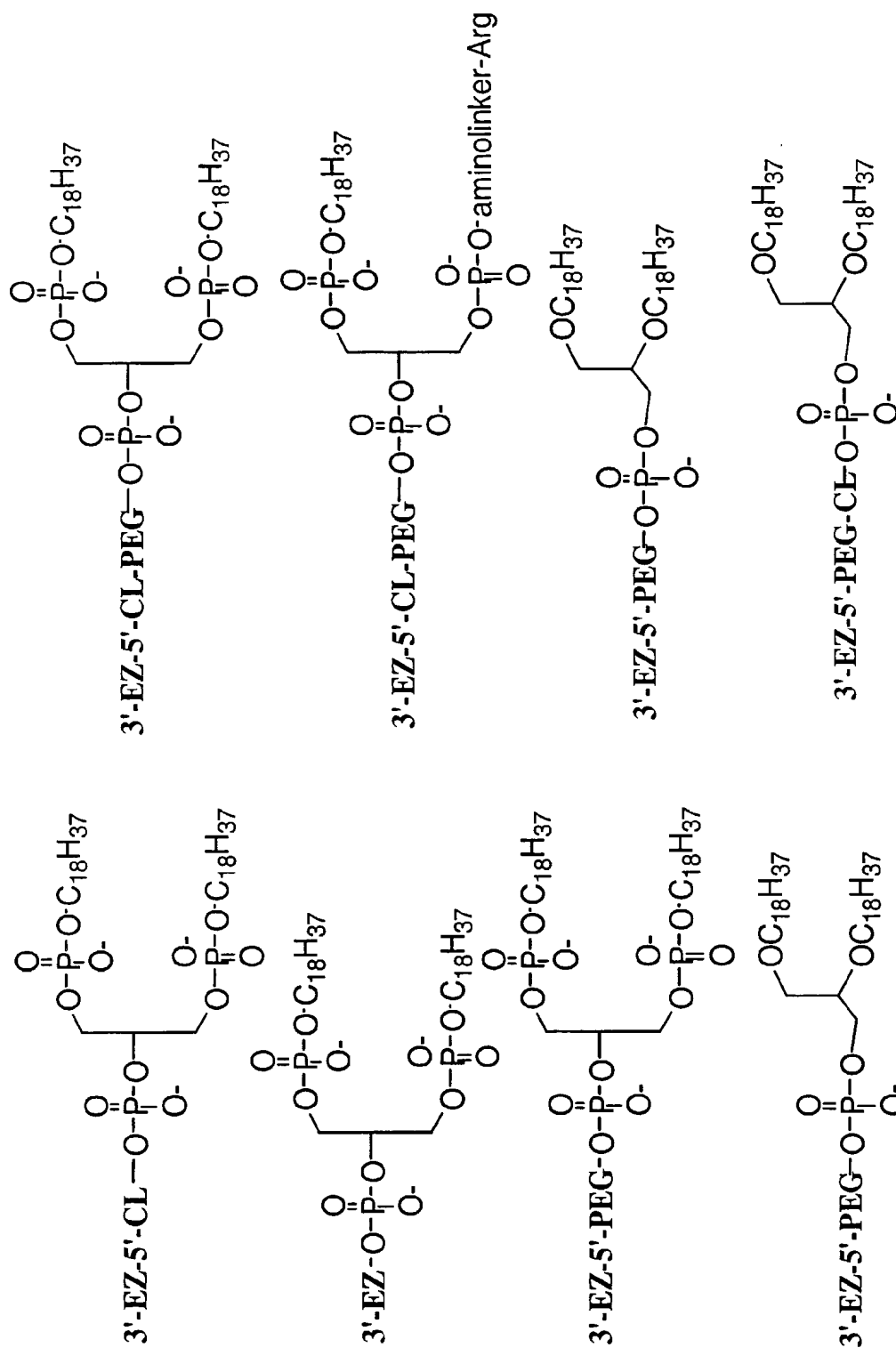
PEG=polyethylene glycol
CL=cleavable linker (e.g. A-dT, C-dT)
EZ= Enzymatic Nucleic Acid Molecule

Figure 19: Angiozyme Phospholipid Conjugate

Figure 20: Peptide PEG Conjugate
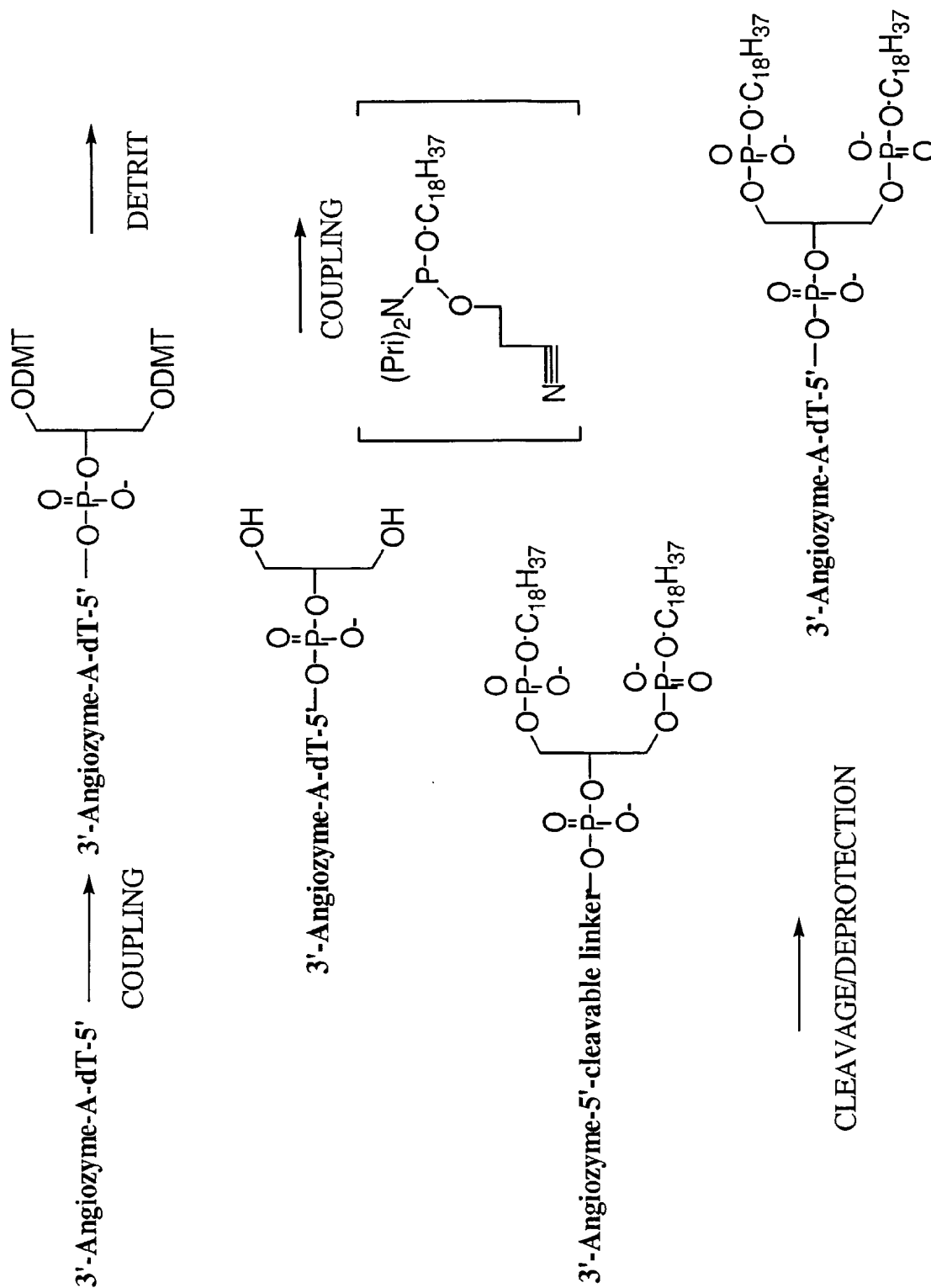
A = Adenosine

Figure 21:   40-KDa PEG-Angiozyme vs Angiozyme
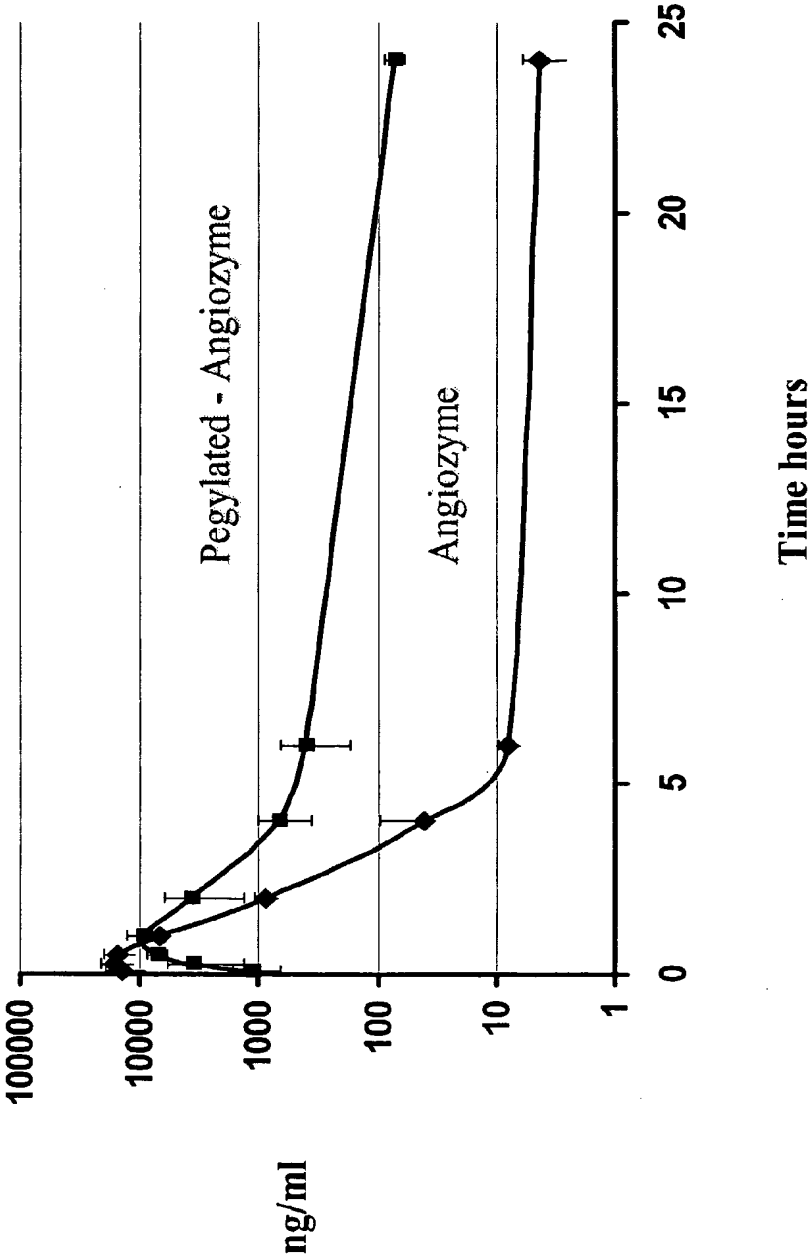
Figure 22: *Phospholipid-Angiozyme vs Angiozyme*

Figure 23: Oligonucleotide-NAcGalactosamine post-synthetic coupling

OLIGO—LINKER—(NH₂)₃   OLIGO—LINKER $\left(\text{—NHCO(CH}_2)_{11}\cdots \right)_3$

(Sulfo-NHS, EDC, 1M HEPES, pH 8, DMSO)

FOR EXAMPLE: OLIGO-LINKER =

Where n is an integer from 1 to 20
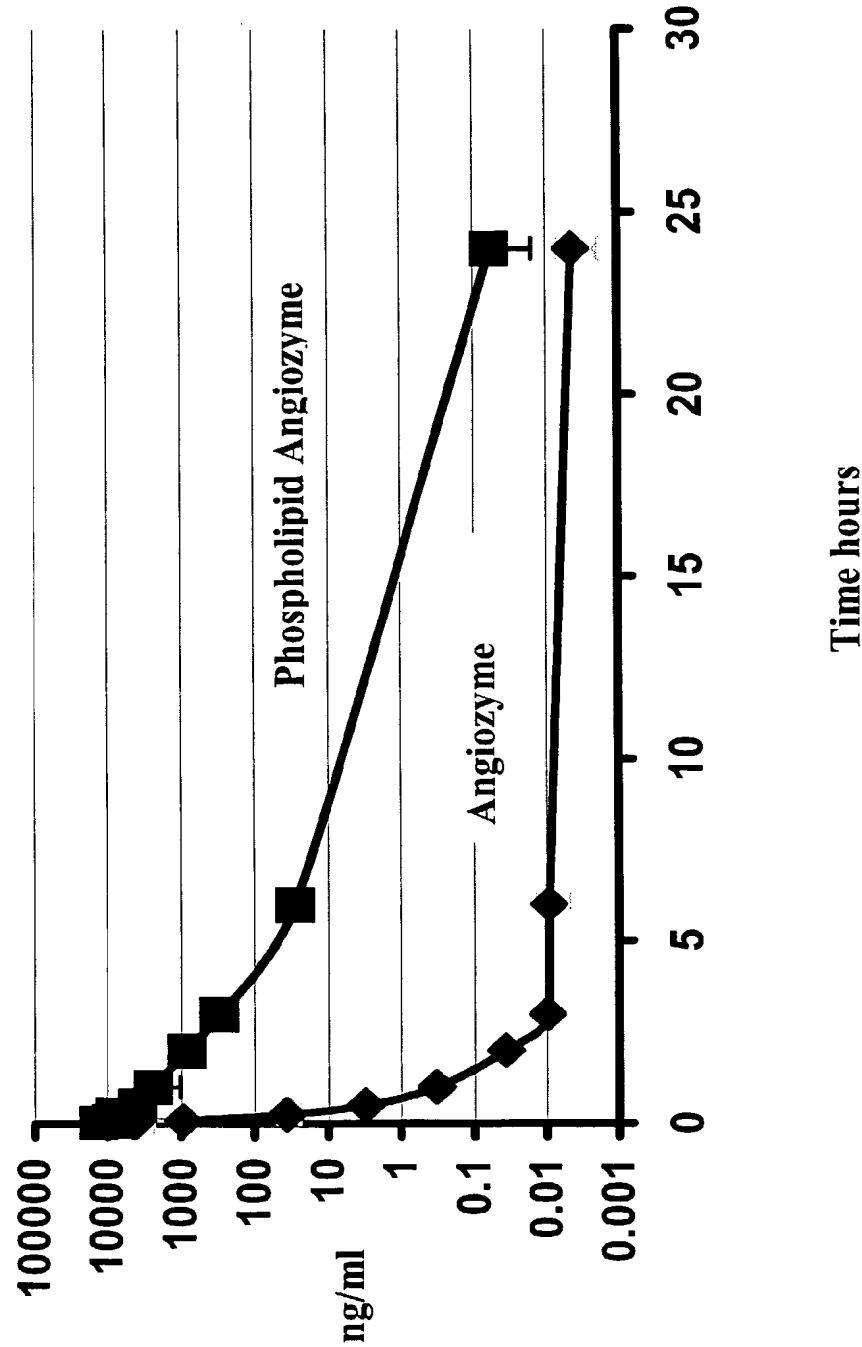
Figure 24a: Protein with cleavable linker

B_1 = T, U
B_2 = C, U, A
B = A, G, C, U, T, or H
R_1= O, S
R_2= O, S, N, alkyl etc.
R_3= O, S, N, C
R_4= O, S, N, C
R_5= H, OH, O-alkyl etc.
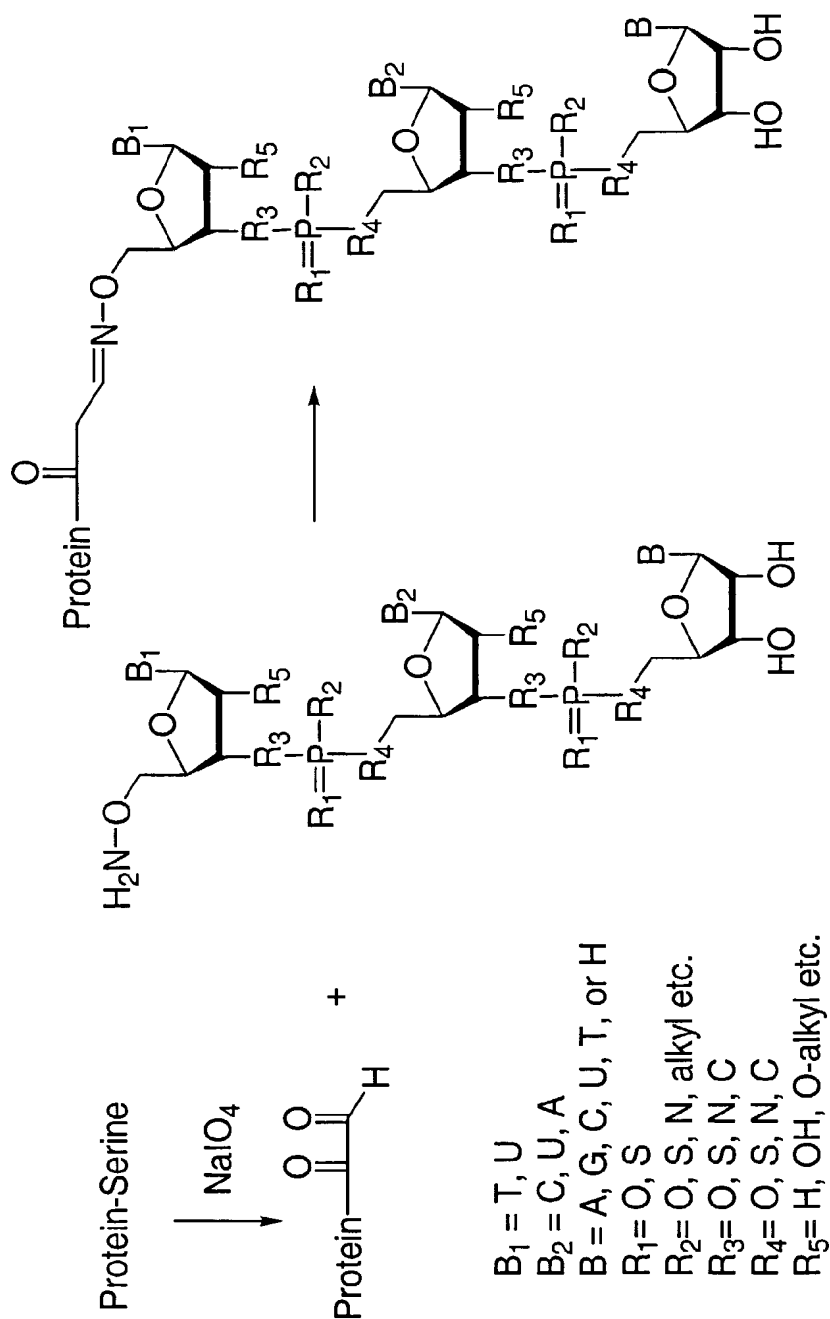
Figure 24b: Protein cleavable linker PEG Conjugate
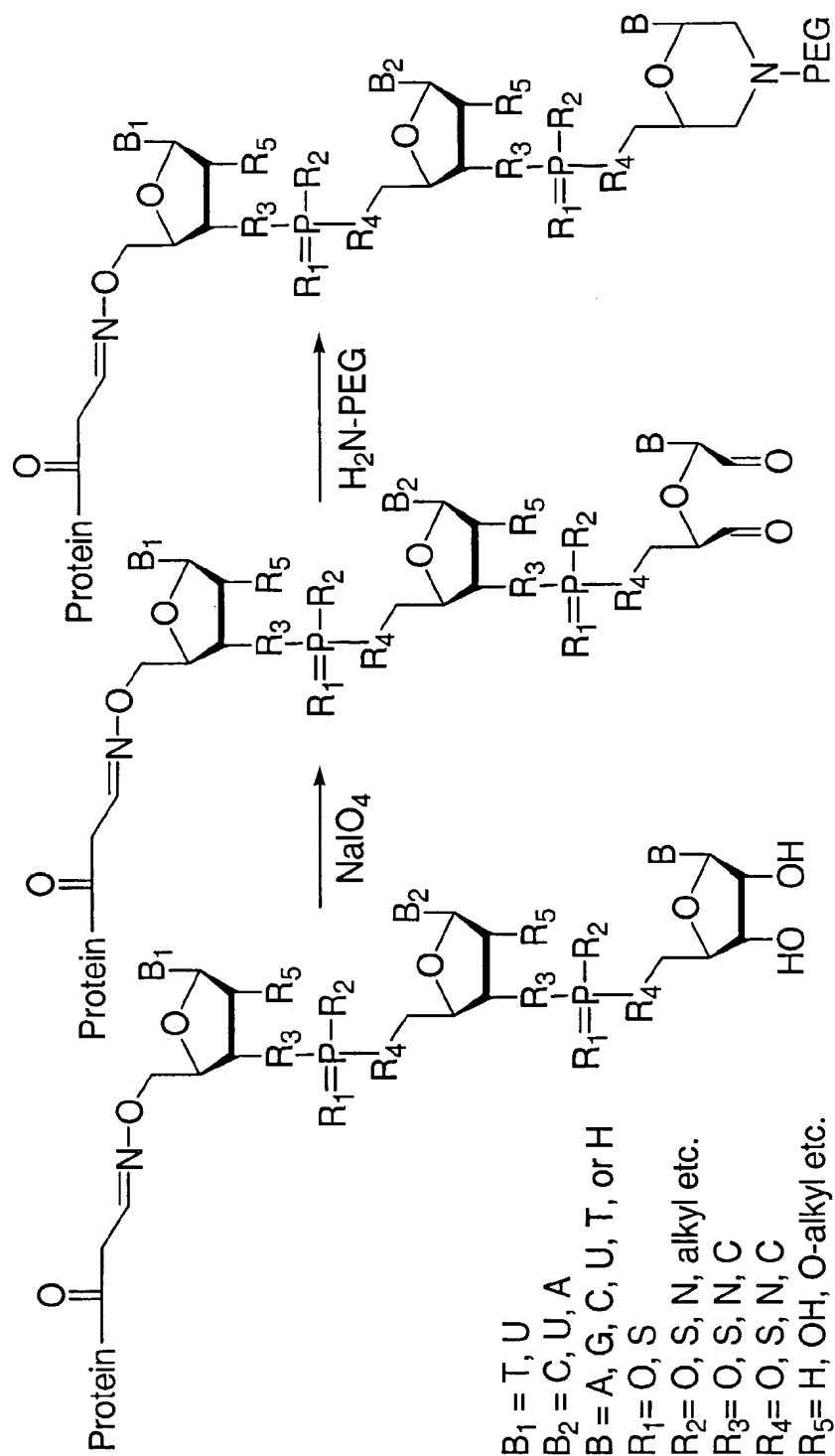
Figure 25: Protein PEG conjugate with cleavable linker

B₁ = T, U
B₂ = C, U, A
R₁ = O, S
R₂ = O, S, N, alkyl etc.
R₃ = O, S, N, C
R₄ = O, S, N, C
R₅ = H, OH, O-alkyl etc.
NMI = N-METHYLIMIDAZOLE OR OTHER ACTIVATOR
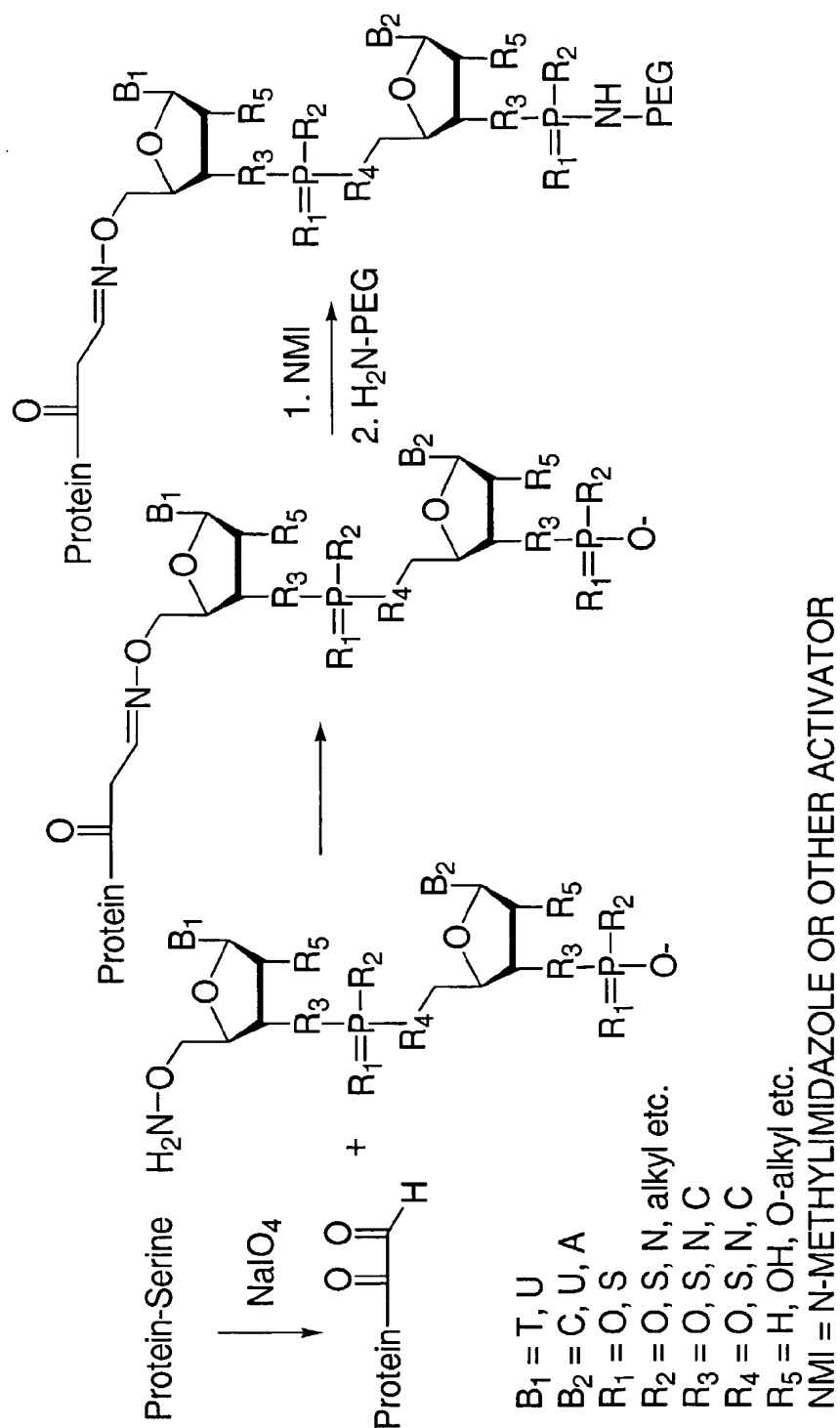
Figure 26a: PEG with cleavable linker
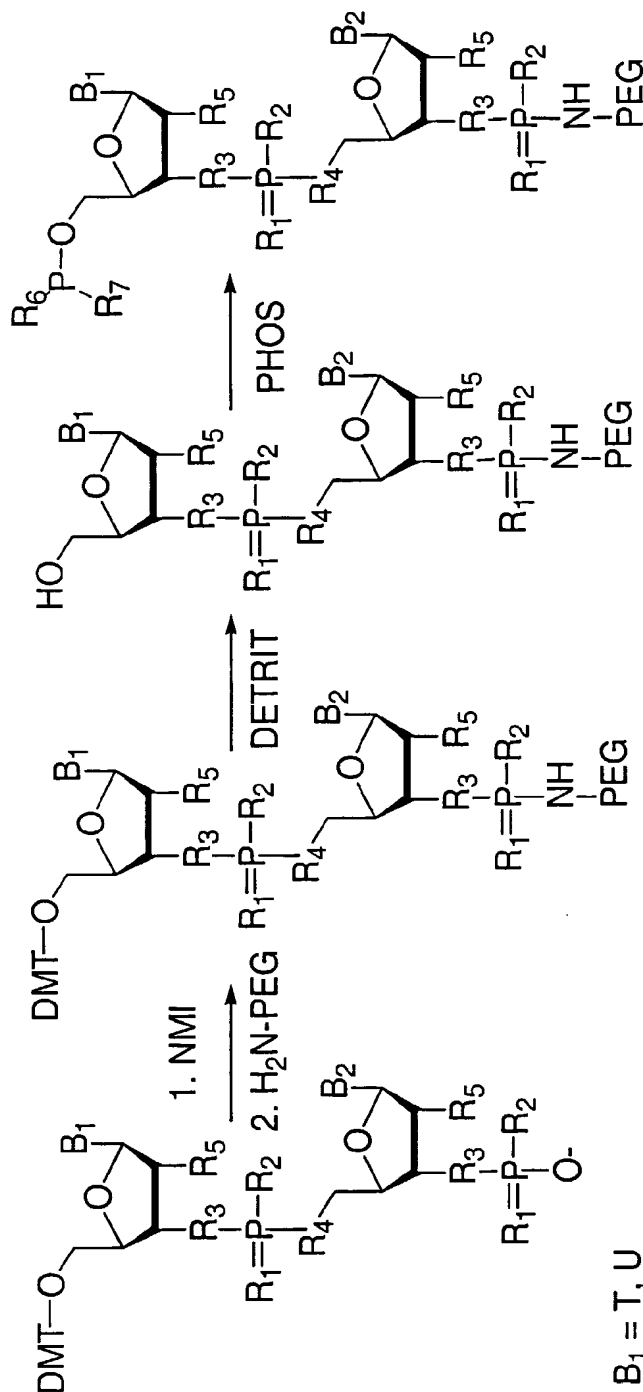
B₁ = T, U
B₂ = C, U, A
R₁ = O, S
R₂ = O, S, N, alkyl etc.
R₃ = O, S, N, C
R₄ = O, S, N, C
R₅ = H, OH, O-alkyl etc.

R₆ = substituted nitrogen, eg. N(iPr)₂
R₇ = O or S alkyl, cyanoalkyl, eg. cyanoethyl
PHOS = PHOSPHITYLATION
DETRIT = DETRITYLATION
NMI = N-METHYLIMIDAZOLE OR OTHER ACTIVATOR

Figure 26b: Protein PEG conjugate with cleavable linker

1. COUPLE
2. OXIDIZE
3. DEPROTECT

+ H₂N-PROTEIN

B_1 = T, U
B_2 = C, U, A
R_1 = O, S
R_2 = O, S, N, alkyl etc.
R_3 = O, S, N, C
R_4 = O, S, N, C
R_5 = H, OH, O-alkyl etc.
R_6 = substituted nitrogen, eg. N(iPr)₂
R_7 = O or S alkyl, cyanoalkyl, eg. cyanoethyl
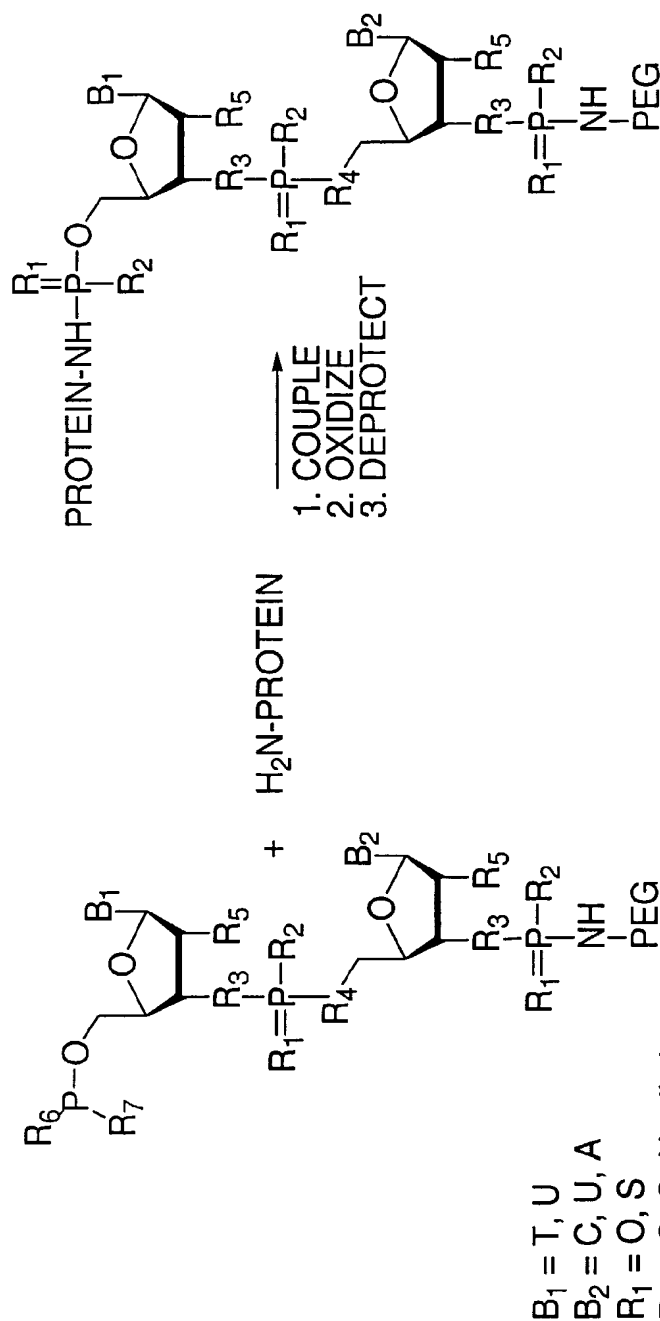
Figure 27: Peptide or Protein/Phospholipid Conjugates
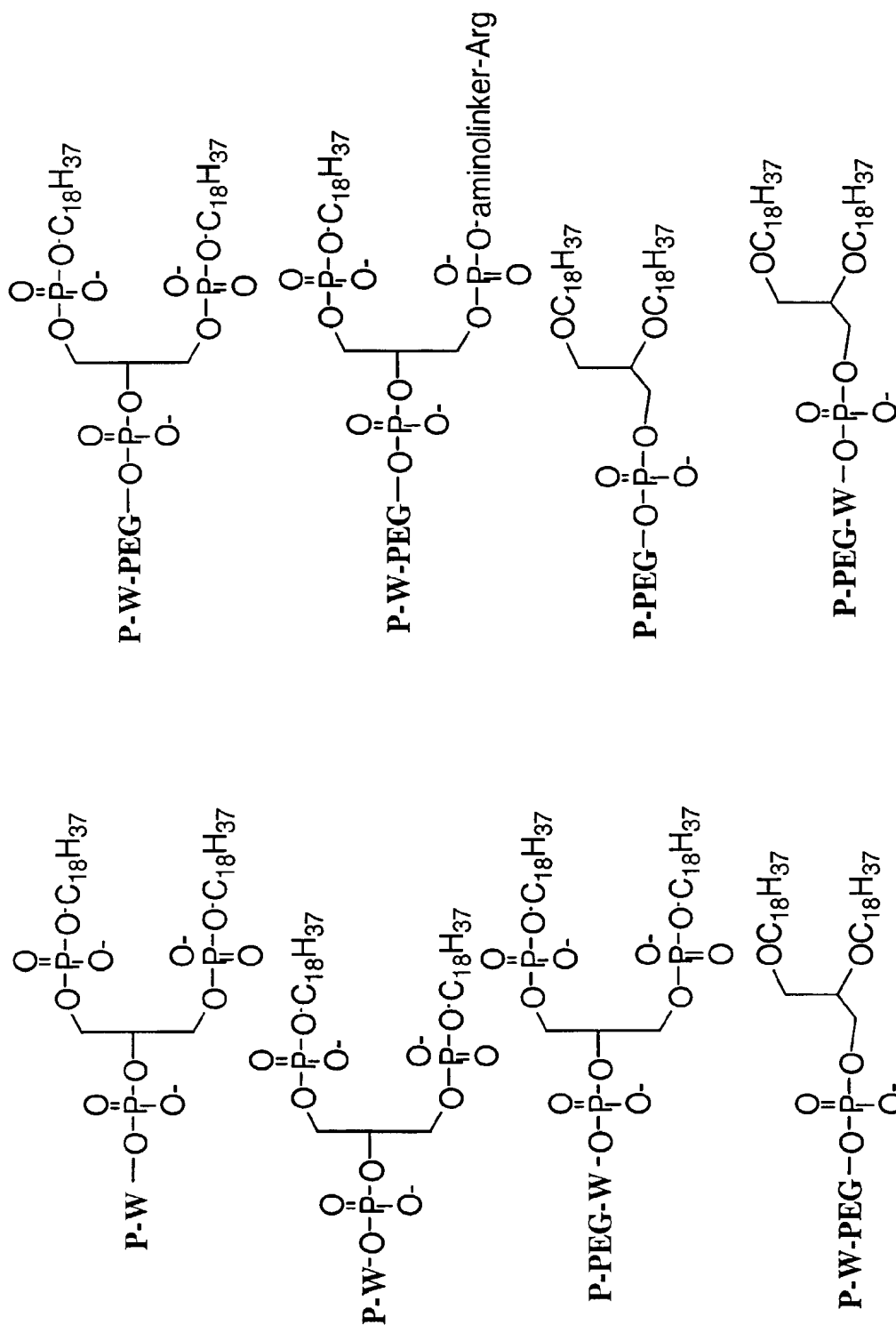
PEG=polyethylene glycol
W=cleavable linker (e.g. A-dT, C-dT)
P= Peptide/Protein

Figure 28: Conjugation of targeting ligands to a peptide or protein

N-acetyl-D-galactosamine conjugate

W = cleavable linker (eg. A-dT, C-dT dimer)

Figure 29: Protein/PEG conjugate with cleavable linker

$B_1 = T, U$
$B_2 = C, U, A$
$R_1 = O, S$
$R_2 = O, S, N$, alkyl etc.
$R_3 = O, S, N, C$
$R_4 = O, S, N, C$
$R_5 = H, OH, O$-alkyl etc.

$R_6 =$ substituted nitrogen, eg. $N(iPr)_2$
$R_7 = O$ or S alkyl, cyanoalkyl, eg. cyanoethyl
PHOS = PHOSPHITYLATION

(54) Title: RNA SEQUENCE-SPECIFIC MEDIATORS OF RNA INTERFERENCE

(57) Abstract: The present invention relates to a Drosophila in vitro system which was used to demonstrate that dsRNA is processed to RNA segments 21-23 nucleotides (nt) in length. Furthermore, when these 21-23 nt fragments are purified and added back to Drosophila extracts, they mediate RNA interference in the absence of long dsRNA. Thus , these 21-23 nt fragments are the sequence-specific mediators of RNA degradation. A molecular signal, which may be their specific length, must be present in these 21-23 nt fragments to recruit cellular factors involved in RNAi. This present invention encompasses these 21-23 nt fragments and their use for specifically inactivating gene function. The use of these fragments (or chemically synthesized oligonucleotides of the same or similar nature) enables the targeting of specific mRNAs for degradation in mammalian cells, where the use of long dsRNAs to elicit RNAi is usually not practical, presumably because of the deleterious effects of the interferon response. This specific targeting of a particular gene function is useful in functional genomic and therapeutic applications.

RNA Sequence-Specific Mediators of RNA Interference

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 60/265,232, filed January 31, 2001 and U.S. Provisional Application No.
5   60/193,594, filed March 30, 2000, and claims priority under 35 U.S.C. §119 to European Application No. 00 126 325.0 filed December 1, 2000. The entire teachings of the above applications are incorporated herein by reference.

GOVERNMENT SUPPORT

Work described herein was funded in part by grants from the National
10   Institutes of Health through a United States Public Health Service MERIT award (Grant No. RO1-GM34277) from the National Institutes of Health. The United States government has certain rights in the invention.

BACKGROUND OF THE INVENTION

RNA interference or "RNAi" is a term initially coined by Fire and
15   co-workers to describe the observation that double-stranded RNA (dsRNA) can block gene expression when it is introduced into worms (Fire et al. (1998) Nature 391, 806-811). dsRNA directs gene-specific, post-transcriptional silencing in many organisms, including vertebrates, and has provided a new tool for studying gene function. RNAi involves mRNA degradation, but many of the biochemical
20   mechanisms underlying this interference are unknown. The recapitulation of the essential features of RNAi in vitro is needed for a biochemical analysis of the phenomenon.

SUMMARY OF THE INVENTION

Described herein is gene-specific, dsRNA-mediated interference in a
25   cell-free system derived from syncytial blastoderm Drosophila embryos. The in vitro system complements genetic approaches to dissecting the molecular basis of

-2-

RNAi. As described herein, the molecular mechanisms underlying RNAi were examined using the Drosophila in vitro system. Results showed that RNAi is ATP-dependent yet uncoupled from mRNA translation. That is, protein synthesis is not required for RNAi in vitro. In the RNAi reaction, both strands (sense and

5   antisense) of the dsRNA are processed to small RNA fragments or segments of from about 21 to about 23 nucleotides (nt) in length (RNAs with mobility in sequencing gels that correspond to markers that are 21-23 nt in length, optionally referred to as 21-23 nt RNA). Processing of the dsRNA to the small RNA fragments does not require the targeted mRNA, which demonstrates that the small RNA species is

10 generated by processing of the dsRNA and not as a product of dsRNA-targeted mRNA degradation. The mRNA is cleaved only within the region of identity with the dsRNA. Cleavage occurs at sites 21-23 nucleotides apart, the same interval observed for the dsRNA itself, suggesting that the 21-23 nucleotide fragments from the dsRNA are guiding mRNA cleavage. That purified 21-23 nt RNAs mediate

15 RNAi confirms that these fragments are guiding mRNA cleavage.

Accordingly, the present invention relates to isolated RNA molecules (double- stranded; single-stranded) of from about 21 to about 23 nucleotides which mediate RNAi. That is, the isolated RNAs of the present invention mediate degradation of mRNA of a gene to which the mRNA corresponds (mediate

20 degradation of mRNA that is the transcriptional product of the gene, which is also referred to as a target gene). For convenience, such mRNA is also referred to herein as mRNA to be degraded. As used herein, the terms RNA, RNA molecule(s), RNA segment(s) and RNA fragment(s) are used interchangeably to refer to RNA that mediates RNA interference. These terms include double-stranded RNA,

25 single-stranded RNA, isolated RNA (partially purified RNA, essentially pure RNA, synthetic RNA, recombinantly produced RNA), as well as altered RNA that differs from naturally occurring RNA by the addition, deletion, substitution and/or alteration of one or more nucleotides. Such alterations can include addition of non-nucleotide material, such as to the end(s) of the 21-23 nt RNA or internally (at

30 one or more nucleotides of the RNA). Nucleotides in the RNA molecules of the present invention can also comprise non-standard nucleotides, including

-3-

non-naturally occurring nucleotides or deoxyribonucleotides. Collectively, all such altered RNAs are referred to as analogs or analogs of naturally-occurring RNA. RNA of 21-23 nucleotides of the present invention need only be sufficiently similar to natural RNA that it has the ability to mediate (mediates) RNAi. As used herein the phrase "mediates RNAi" refers to (indicates) the ability to distinguish which RNAs are to be degraded by the RNAi machinery or process. RNA that mediates RNAi interacts with the RNAi machinery such that it directs the machinery to degrade particular mRNAs. In one embodiment, the present invention relates to RNA molecules of about 21 to about 23 nucleotides that direct cleavage of specific mRNA to which their sequence corresponds. It is not necessary that there be perfect correspondence of the sequences, but the correspondence must be sufficient to enable the RNA to direct RNAi cleavage of the target mRNA. In a particular embodiment, the 21-23 nt RNA molecules of the present invention comprise a 3' hydroxyl group.

The present invention also relates to methods of producing RNA molecules of about 21 to about 23 nucleotides with the ability to mediate RNAi cleavage. In one embodiment, the Drosophila in vitro system is used. In this embodiment, dsRNA is combined with a soluble extract derived from Drosophila embryo, thereby producing a combination. The combination is maintained under conditions in which the dsRNA is processed to RNA molecules of about 21 to about 23 nucleotides. In another embodiment, the Drosophila in vitro system is used to obtain RNA sequences of about 21 to about 23 nucleotides which mediate RNA interference of the mRNA of a particular gene (e.g., oncogene, viral gene). In this embodiment, double-stranded RNA that corresponds to a sequence of the gene to be targeted is combined with a soluble extract derived from Drosophila embryo, thereby producing a combination. The combination is maintained under conditions in which the double-stranded RNA is processed to RNA of about 21 to about 23 nucleotides in length. As shown herein, 21- 23 nt RNA mediates RNAi of the mRNA of the targeted gene (the gene whose mRNA is to be degraded). The method of obtaining 21-23 nt RNAs using the Drosophila in vitro system can further comprise isolating the RNA sequence from the combination.

-4-

The present invention also relates to 21-23 nt RNA produced by the methods of the present invention, as well as to 21-23 nt RNAs, produced by other methods, such as chemical synthesis or recombinant DNA techniques, that have the same or substantially the same sequences as naturally-occurring RNAs that mediate RNAi,

5   such as those produced by the methods of the present invention. All of these are referred to as 21-23 nt RNAs that mediate RNA interference. As used herein, the term isolated RNA includes RNA obtained by any means, including processing or cleavage of dsRNA as described herein; production by chemical synthetic methods; and production by recombinant DNA techniques. The invention further relates to

10   uses of the 21-23 nt RNAs, such as for therapeutic or prophylactic treatment and compositions comprising 21-23 nt RNAs that mediate RNAi, such as pharmaceutical compositions comprising 21-23 nt RNAs and an appropriate carrier (e.g., a buffer or water).

The present invention also relates to a method of mediating RNA

15   interference of mRNA of a gene in a cell or organism (e.g., mammal such as a mouse or a human). In one embodiment, RNA of about 21 to about 23 nt which targets the mRNA to be degraded is introduced into the cell or organism. The cell or organism is maintained under conditions under which degradation of the mRNA occurs, thereby mediating RNA interference of the mRNA of the gene in the cell or

20   organism. The cell or organism can be one in which RNAi occurs as the cell or organism is obtained or a cell or organism can be one that has been modified so that RNAi occurs (e.g., by addition of components obtained from a cell or cell extract that mediate RNAi or activation of endogenous components). As used herein, the term "cell or organism in which RNAi occurs" includes both a cell or organism in

25   which RNAi occurs as the cell or organism is obtained, or a cell or organism that has been modified so that RNAi occurs. In another embodiment, the method of mediating RNA interference of a gene in a cell comprises combining double-stranded RNA that corresponds to a sequence of the gene with a soluble extract derived from Drosophila embryo, thereby producing a combination. The

30   combination is maintained under conditions in which the double-stranded RNA is processed to RNAs of about 21 to about 23 nucleotides. 21 to 23 nt RNA is then

-5-

isolated and introduced into the cell or organism. The cell or organism is maintained under conditions in which degradation of mRNA of the gene occurs, thereby mediating RNA interference of the gene in the cell or organism. As described for the previous embodiment, the cell or organism is one in which RNAi occurs naturally

5   (in the cell or organism as obtained) or has been modified in such a manner that RNAi occurs.   21 to 23 nt RNAs can also be produced by other methods, such as chemical synthetic methods or recombinant DNA techniques.

The present invention also relates to biochemical components of a cell, such as a Drosophila cell, that process dsRNA to RNA of about 21 to about 23

10   nucleotides. In addition, biochemical components of a cell that are involved in targeting of mRNA by RNA of about 21 to about 23 nucleotides are the subject of the present invention. In both embodiments, the biochemical components can be obtained from a cell in which they occur or can be produced by other methods, such as chemical synthesis or recombinant DNA methods.   As used herein, the

15   term "isolated" includes materials (e.g., biochemical components, RNA) obtained from a source in which they occur and materials produced by methods such as chemical synthesis or recombinant nucleic acid (DNA, RNA) methods.

The present invention also relates to a method for knocking down (partially or completely) the targeted gene, thus providing an alternative to presently available

20   methods of knocking down (or out) a gene or genes. This method of knocking down gene expression can be used therapeutically or for research (e.g., to generate models of disease states, to examine the function of a gene, to assess whether an agent acts on a gene, to validate targets for drug discovery). In those instances in which gene function is eliminated, the resulting cell or organism can also be referred to as a

25   knockout. One embodiment of the method of producing knockdown cells and organisms comprises introducing into a cell or organism in which a gene (referred to as a targeted gene) is to be knocked down, RNA of about 21 to about 23 nt that targets the gene and maintaining the resulting cell or organism under conditions under which RNAi occurs, resulting in degradation of the mRNA of the targeted

30   gene, thereby producing knockdown cells or organisms. Knockdown cells and organisms produced by the present method are also the subject of this invention.

-6-

The present invention also relates to a method of examining or assessing the function of a gene in a cell or organism. In one embodiment, RNA of about 21 to about 23 nt which targets mRNA of the gene for degradation is introduced into a cell or organism in which RNAi occurs. The cell or organism is referred to as a test cell

5   or organism. The test cell or organism is maintained under conditions under which degradation of mRNA of the gene occurs. The phenotype of the test cell or organism is then observed and compared to that of an appropriate control cell or organism, such as a corresponding cell or organism that is treated in the same manner except that the targeted (specific) gene is not targeted. A 21 to 23 nt RNA

10   that does not target the mRNA for degradation can be introduced into the control cell or organism in place of the RNA introduced into the test cell or organism, although it is not necessary to do so. A difference between the phenotypes of the test and control cells or organisms provides information about the function of the degraded mRNA. In another embodiment, double-stranded RNA that corresponds to

15   a sequence of the gene is combined with a soluble extract that mediates RNAi, such as the soluble extract derived from Drosophila embryo described herein, under conditions in which the double-stranded RNA is processed to generate RNA of about 21 to about 23 nucleotides. The RNA of about 21 to about 23 nucleotides is isolated and then introduced into a cell or organism in which RNAi occurs (test cell

20   or test organism). The test cell or test organism is maintained under conditions under which degradation of the mRNA occurs. The phenotype of the test cell or organism is then observed and compared to that of an appropriate control, such as a corresponding cell or organism that is treated in the same manner as the test cell or organism except that the targeted gene is not targeted. A difference between the

25   phenotypes of the test and control cells or organisms provides information about the function of the targeted gene. The information provided may be sufficient to identify (define) the function of the gene or may be used in conjunction with information obtained from other assays or analyses to do so.

Also the subject of the present invention is a method of validating whether an agent acts on a gene. In this method, RNA of from about 21 to about 23

30   nucleotides that targets the mRNA to be degraded is introduced into a cell or

organism in which RNAi occurs. The cell or organism (which contains the introduced RNA) is maintained under conditions under which degradation of mRNA occurs, and the agent is introduced into the cell or organism. Whether the agent has an effect on the cell or organism is determined; if the agent has no effect

5   on the cell or organism, then the agent acts on the gene.

The present invention also relates to a method of validating whether a gene product is a target for drug discovery or development. RNA of from about 21 to about 23 nucleotides that targets the mRNA that corresponds to the gene for degradation is introduced into a cell or organism. The cell or organism is

10   maintained under conditions in which degradation of the mRNA occurs, resulting in decreased expression of the gene. Whether decreased expression of the gene has an effect on the cell or organism is determined, wherein if decreased expression of the gene has an effect, then the gene product is a target for drug discovery or development.

15   The present invention also encompasses a method of treating a disease or condition associated with the presence of a protein in an individual comprising administering to the individual RNA of from about 21 to about 23 nucleotides which targets the mRNA of the protein (the mRNA that encodes the protein) for degradation. As a result, the protein is not produced or is not produced to the extent

20   it would be in the absence of the treatment.

Also encompassed by the present invention is a gene identified by the sequencing of endogenous 21 to 23 nucleotide RNA molecules that mediate RNA interference.

Also encompassed by the present invention is a method of identifying target

25   sites within an mRNA that are particularly suitable for RNAi as well as a method of assessing the ability of 21-23 nt RNAs to mediate RNAi.

BRIEF DESCRIPTION OF THE DRAWINGS

The file of this patent contains at least one drawing executed in color. Copies of this patent with color drawing(s) will be provided by the Patent and

30   Trademark Office upon request and payment of the necessary fee.

-8-

Figure 1 is a schematic representation of reporter mRNAs and dsRNAs Rr-Luc and Pp-Luc. Lengths and positions of the ssRNA, asRNA, and dsRNAs are shown as black bars relative to the Rr-Luc and Pp-Luc reporter mRNA sequences. Black rectangles indicate the two unrelated luciferase coding sequences, lines

5   correspond to the 5' and 3' untranslated regions of the mRNAs.

Figure 2A is a graph of the ratio of luciferase activities after targeting 50 pM Pp- Luc mRNA with 10 nM ssRNA, asRNA, or dsRNA from the 505 bp segment of the Pp- Luc gene showing gene-specific interference by dsRNA in vitro. The data are the average values of seven trials ± standard deviation. Four independently

10   prepared lysates were used. Luciferase activity was normalized to the buffer control; a ratio equal to one indicates no gene-specific interference.

Figure 2B is a graph of the ratio of luciferase activities after targeting 50 pM Rr- Luc mRNA with 10 nM ssRNA, asRNA, or dsRNA from the 501 bp segment of the Rr- Luc gene showing gene-specific interference by dsRNA in vitro. The data

15   are the average values of six trials ± standard deviation. A Rr-Luc/Pp-Luc ratio equal to one indicates no gene-specific interference.

Figure 3A is a schematic representation of the experimental strategy used to show that incubation in the Drosophila embryo lysate potentiates dsRNA for gene-specific interference. The same dsRNAs used in Figure 2 (or buffer) was serially

20   preincubated using two-fold dilutions in six successive reactions with Drosophila embryo lysate, then tested for its capacity to block mRNA expression. As a control, the same amount of dsRNA (10 nM) or buffer was diluted directly in buffer and incubated with Pp-Luc and Rr-Luc mRNAs and lysate.

Figure 3B is a graph of potentiation when targeting Pp-Luc mRNA. Black

25   columns indicate the dsRNA or the buffer was serially preincubated; white columns correspond to a direct 32-fold dilution of the dsRNA. Values were normalized to those of the buffer controls.

Figure 3C is a graph of potentiation when targeting Rr-Luc mRNA. The corresponding buffer control is shown in Figure 3B.

30   Figure 4 is a graph showing effect of competitor dsRNA on gene-specific interference. Increasing concentrations of nanos dsRNA (508 bp) were added to

reactions containing 5 nM dsRNA (the same dsRNAs used in Figures 2A and 2B) targeting Pp-Luc mRNA (black columns, left axis) or Rr-Luc mRNA (white columns, right axis). Each reaction contained both a target mRNA (Pp-Luc for the black columns, Rr-Luc for the white) and an unrelated control mRNA (Rr-Luc for
5   the black columns, Pp-Luc for the white). Values were normalized to the buffer control (not shown). The reactions were incubated under standard conditions (see Methods).

Figure 5A is a graph showing the effect of dsRNA on mRNA stability. Circles, Pp-Luc mRNA; squares, Rr-Luc mRNA; filled symbols, buffer incubation;
10   open symbols, incubation with Pp-dsRNA.

Figure 5B is a graph showing the stability of Rr-Luc mRNA incubated with Rr- dsRNA or Pp-dsRNA. Filled squares, buffer; open squares, Pp-dsRNA (10 nM); open circles, Rr-dsRNA (10 nM).

Figure 5C is a graph showing the dependence on dsRNA length. The
15 . stability of the Pp-Luc mRNA was assessed after incubation in lysate in the presence of buffer or dsRNAs of different lengths. Filled squares, buffer; open circles, 49 bp dsRNA (10 nM); open inverted triangles, 149 bp dsRNA (10 nM); open triangles, 505 bp dsRNA (10 nM); open diamonds, 997 bp dsRNA (10 nM). Reactions were incubated under standard conditions (see Methods).

20   Figure 6 is a graph showing that RNAi Requires ATP. Creatine kinase (CK) uses creatine phosphate (CP) to regenerate ATP. Circles, +ATP, +CP, +CK; squares, -ATP, +CP, +CK; triangles, -ATP, -CP, +CK; inverted triangles, -ATP, +CP, -CK.

Figure 7A is a graph of protein synthesis, as reflected by luciferase activity
25   produced after incubation of Rr-luc mRNA in the in vitro RNAi reaction for 1 hour, in the presence of the protein synthesis inhibitors anisomycin, cycloheximide, or chloramphenicol, relative to a reaction without any inhibitor showing that RNAi does not require mRNA translation.

Figure 7B is a graph showing translation of 7-methyl-guanosine- and
30   adenosine- capped Pp-luc mRNAs (circles and squares, respectively) in the RNAi

-10-

reaction in the absence of dsRNA, as measured by luciferase activity produced in a one-hour incubation.

Figure 7C is a graph showing incubation in an RNAi reaction of uniformly ^{32}P- radiolabeled 7-methyl-guanosine-capped Pp-luc mRNA (circles) and

5   adenosine-capped Pp-luc mRNA (squares), in the presence (open symbols) and absence (filled symbols) of 505 bp Pp-luc dsRNA.

Figure 8A is a graph of the of the denaturing agarose-gel analysis of Pp-luc mRNA incubated in a standard RNAi reaction with buffer, 505 nt Pp-asRNA, or 505 bp Pp-dsRNA for the times indicated showing that asRNA causes a small amount of

10   RNAi in vitro.

Figure 8B is a graph of the of the denaturing agarose-gel analysis of Rr-luc mRNA incubated in a standard RNAi reaction with buffer, 505 nt Pp-asRNA, or 505 bp Pp-dsRNA for the times indicated showing that asRNA causes a small amount of RNAi in vitro.

15   Figure 9 is a schematic of the positions of the three dsRNAs, 'A,' 'B,' and 'C,' relative to the Rr-luc mRNA.

Figure10 indicates the cleavage sites mapped onto the first 267 nt of the Rr-luc mRNA (SEQ ID NO: 1). The blue bar below the sequence indicates the position of dsRNA 'C,' and blue circles indicate the position of cleavage sites caused

20   by this dsRNA. The green bar denotes the position of dsRNA 'B,' and green circles, the cleavage sites. The magenta bar indicates the position of dsRNA 'A,' and magenta circles, the cleavages. An exceptional cleavage within a run of 7 uracils is marked with a red arrowhead.

Figure 11 is a proposed model for RNAi. RNAi is envisioned to begin with

25   cleavage of the dsRNA to 21-23 nt products by a dsRNA-specific nuclease, perhaps in a multiprotein complex. These short dsRNAs might then be dissociated by an ATP- dependent helicase, possibly a component of the initial complex, to 21-23 nt asRNAs that could then target the mRNA for cleavage. The short asRNAs are imagined to remain associated with the RNAi-specific proteins (circles) that were

30   originally bound by the full-length dsRNA, thus explaining the inefficiency of

-11-

asRNA to trigger RNAi in vivo and in vitro. Finally, a nuclease (triangles) would cleave the mRNA.

Figure 12 is a bar graph showing sequence-specific gene silencing by 21-23 nt fragments. Ratio of luciferase activity after targeting of Pp-Luc and Rr-Luc

5   mRNA by 5 nM Pp-Luc or Rr-Luc dsRNA (500 bp) or 21-23 nt fragments isolated from a previous incubation of the respective dsRNA in Drosophila lysate. The amount of isolated 21-23 mers present in the incubation reaction correspond to approximately the same amount of 21-23 mers generated during an incubation reaction with 5 nM 500 bp dsRNA. The data are average values of 3 trials and the

10   standard deviation is given by error bars. Luciferase activity was normalized to the buffer control.

Figure 13A illustrates the purification of RNA fragments on a Superdex HR 200 10/30 gel filtration column (Pharmacia) using the method described in Example 4. dsRNA was 32P-labeled, and the radioactivity recovered in each column fraction

15   is graphed. The fractions were also analyzed by denaturing gel electrophoresis (inset).

Figure 13B demonstrates the ability of the Rr-luciferase RNA, after incubation in the Drosophila lysate and fractionation as in Fig. 13A, to mediate sequence-specific interference with the expression of a Rr-luciferase target mRNA.

20   One microliter of each resuspended fraction was tested in a 10 microliter in vitro RNAi reaction (see Example 1). This procedure yields a concentration of RNA in the standard in vitro RNAi reaction that is approximately equal to the concentration of that RNA species in the original reaction prior to loading on the column. Relative luminescence per second has been normalized to the average value of the two buffer

25   controls.

Figure 13C is the specificity control for Fig 13B. It demonstrates that the fractionated RNA of Fig 13B does not efficiently mediate sequence-specific interference with the expression of a Pp-luciferase mRNA. Assays are as in Fig 13B.

30   Figures 14A and 14B are schematic representations of reporter constructs and siRNA duplexes. Figure 14A illustrates the firefly (Pp-luc) and sea pansy (Rr-

-12-

luc) luciferase reporter gene regions from plasmids pGL2-Control, pGL3-Control, and pRL-TK (Promega). SV40 regulatory elements, the HSV thymidine kinase promoter, and two introns (lines) are indicated. The sequence of GL3 luciferase is 95% identical to GL2, but RL is completely unrelated to both. Luciferase expression
5   from pGL2 is approximately 10-fold lower than from pGL3 in transfected mammalian cells. The region targeted by the siRNA duplexes is indicated as black bar below the coding region of the luciferase genes. Figure 14B shows the sense (top) and antisense (bottom) sequences of the siRNA duplexes targeting GL2 (SEQ ID Nos: 10 and 11), GL3 (SEQ ID Nos: 12 and 13), and RL (SEQ ID Nos: 14 and
10   15) luciferase are shown. The GL2 and GL3 siRNA duplexes differ by only 3 single nucleotide substitutions (boxed in gray). As unspecific control, a duplex with the inverted GL2 sequence, invGL2 (SEQ ID Nos: 16 and 17), was synthesized. The 2 nt 3' overhang of 2'-deoxythymidine is indicated as TT; uGL2 (SEQ ID Nos: 18 and 19) is similar to GL2 siRNA but contains ribo-uridine 3' overhangs.
15   Figures 15A-15J are graphs showing RNA interference by siRNA duplexes. Ratios of target to control luciferase were normalized to a buffer control (bu, black bars); gray bars indicate ratios of *Photinus pyralis* (*Pp*-luc) GL2 or GL3 luciferase to *Renilla reniformis* (*Rr*-luc) RL luciferase (left axis), white bars indicate RL to GL2 or GL3 ratios (right axis). Figures 15A, 15C, 15E, 15G, and 15I show results of
20   experiments performed with the combination of pGL2-Control and pRL-TK reporter plasmids, Figures 15B, 15D, 15F, 15H, and 15J with pGL3-Control and pRL-TK reporter plasmids. The cell line used for the interference experiment is indicated at the top of each plot. The ratios of *Pp*-luc/*Rr*-luc for the buffer control (bu) varied between 0.5 and 10 for pGL2/pRL, and between 0.03 and 1 for pGL3/pRL,
25   respectively, before normalization and between the various cell lines tested. The plotted data were averaged from three independent experiments ± S.D.

       Figures 16A-16F are graphs showing the effects of 21 nt siRNAs, 50 bp, and 500 bp dsRNAs on luciferase expression in HeLa cells. The exact length of the long dsRNAs is indicated below the bars. Figures 16A, 16C, and 16E describe
30   experiments performed with pGL2-Control and pRL-TK reporter plasmids, Figures 16B, 16D, and 16F with pGL3-Control and pRL-TK reporter plasmids. The data

-13-

were averaged from two independent experiments ± S.D. Figures 16A, 16B,
Absolute *Pp*-luc expression, plotted in arbitrary luminescence units. Figure 16C,
16D, Rr-luc expression, plotted in arbitrary luminescence units. Figures 16E, 16F,
Ratios of normalized target to control luciferase. The ratios of luciferase activity for

5   siRNA duplexes were normalized to a buffer control (bu, black bars); the
luminescence ratios for 50 or 500 bp dsRNAs were normalized to the respective
ratios observed for 50 and 500 bp dsRNA from humanized GFP (hG, black bars). It
should be noted, that the overall differences in sequence between the 49 and 484 bp
dsRNAs targeting GL2 and GL3 are not sufficient to confer specificity between GL2

10   and GL3 targets (43 nt uninterrupted identity in 49 bp segment, 239 nt longest
uninterrupted identity in 484 bp segment) (Parrish, S., *et al., Mol. Cell, 6:*1077-1087
(2000)).

DETAILED DESCRIPTION OF THE INVENTION

Double-stranded (dsRNA) directs the sequence-specific degradation of

15   mRNA through a process known as RNA interference (RNAi). The process is
known to occur in a wide variety of organisms, including embryos of mammals and
other vertebrates. Using the Drosophila in vitro system described herein, it has been
demonstrated that dsRNA is processed to RNA segments 21-23 nucleotides (nt) in
length, and furthermore, that when these 21-23 nt fragments are purified and added

20   back to Drosophila extracts, they mediate RNA interference in the absence of longer
dsRNA. Thus, these 21-23 nt fragments are sequence-specific mediators of RNA
degradation. A molecular signal, which may be the specific length of the fragments,
must be present in these 21-23 nt fragments to recruit cellular factors involved in
RNAi. This present invention encompasses these 21-23 nt fragments and their use

25   for specifically inactivating gene function. The use of these fragments (or
recombinantly produced or chemically synthesized oligonucleotides of the same or
similar nature) enables the targeting of specific mRNAs for degradation in
mammalian cells. Use of long dsRNAs in mammalian cells to elicit RNAi is usually
not practical, presumably because of the deleterious effects of the interferon

30   response. Specific targeting of a particular gene function, which is possible with

-14-

21-23 nt fragments of the present invention, is useful in functional genomic and therapeutic applications.

In particular, the present invention relates to RNA molecules of about 21 to about 23 nucleotides that mediate RNAi. In one embodiment, the present invention
5   relates to RNA molecules of about 21 to about 23 nucleotides that direct cleavage of specific mRNA to which they correspond. The 21-23 nt RNA molecules of the present invention can also comprise a 3' hydroxyl group. The 21-23 nt RNA molecules can be single-stranded or double stranded (as two 21-23 nt RNAs); such molecules can be blunt ended or comprise overhanging ends (e.g., 5', 3'). In specific
10   embodiments, the RNA molecule is double stranded and either blunt ended or comprises overhanging ends (as two 21-23 nt RNAs).

In one embodiment, at least one strand of the RNA molecule has a 3' overhang from about 1 to about 6 nucleotides (e.g., pyrimidine nucleotides, purine nucleotides) in length. In other embodiments, the 3' overhang is from about 1 to
15   about 5 nucleotides, from about 1 to about 3 nucleotides and from about 2 to about 4 nucleotides in length. In one embodiment the RNA molecule is double stranded, one strand has a 3' overhang and the other strand can be blunt-ended or have an overhang. In the embodiment in which the RNA molecule is double stranded and both strands comprise an overhang, the length of the overhangs may be the same or
20   different for each strand. In a particular embodiment, the RNA of the present invention comprises 21 nucleotide strands which are paired and which have overhangs of from about 1 to about 3, particularly about 2, nucleotides on both 3' ends of the RNA. In order to further enhance the stability of the RNA of the present invention, the 3' overhangs can be stabilized against degradation. In one
25   embodiment, the RNA is stabilized by including purine nucleotides, such as adenosine or guanosine nucleotides. Alternatively, substitution of pyrimidine nucleotides by modified analogues, e.g., substitution of uridine 2 nucleotide 3' overhangs by 2'-deoxythymidine is tolerated and does not affect the efficiency of RNAi. The absence of a 2' hydroxyl significantly enhances the nuclease resistance
30   of the overhang in tissue culture medium.

-15-

The 21-23 nt RNA molecules of the present invention can be obtained using a number of techniques known to those of skill in the art. For example, the RNA can be chemically synthesized or recombinantly produced using methods known in the art. The 21-23 nt RNAs can also be obtained using the Drosophila in vitro

5   system described herein. Use of the Drosophila in vitro system entails combining dsRNA with a soluble extract derived from Drosophila embryo, thereby producing a combination. The combination is maintained under conditions in which the dsRNA is processed to RNA of about 21 to about 23 nucleotides. The Drosophila in vitro system can also be used to obtain RNA of about 21 to about 23 nucleotides in length

10   which mediates RNA interference of the mRNA of a particular gene (e.g., oncogene, viral gene). In this embodiment, double-stranded RNA that corresponds to a sequence of the gene is combined with a soluble extract derived from Drosophila embryo, thereby producing a combination. The combination is maintained under conditions in which the double- stranded RNA is processed to the RNA of about 21

15   to about 23 nucleotides. As shown herein, 21-23 nt RNA mediates RNAi of the mRNA to be degraded. The present invention also relates to the 21-23 nt RNA molecules produced by the methods described herein.

In one embodiment, the methods described herein are used to identify or obtain 21-23 nt RNA molecules that are useful as sequence-specific mediators of

20   RNA degradation and, thus, for inhibiting mRNAs, such as human mRNAs, that encode products associated with or causative of a disease or an undesirable condition. For example, production of an oncoprotein or viral protein can be inhibited in humans in order to prevent the disease or condition from occurring, limit the extent to which it occurs or reverse it. If the sequence of the gene to be targeted

25   in humans is known, 21-23 nt RNAs can be produced and tested for their ability to mediate RNAi in a cell, such as a human or other primate cell. Those 21-23 nt human RNA molecules shown to mediate RNAi can be tested, if desired, in an appropriate animal model to further assess their in vivo effectiveness. Additional copies of 21-23 nt RNAs shown to mediate RNAi can be produced by the methods

30   described herein.

-16-

The method of obtaining the 21-23 nt RNA sequence using the Drosophila in vitro system can further comprise isolating the RNA sequence from the combination. The 21-23 nt RNA molecules can be isolated using a number of techniques known to those of skill in the art. For example, gel electrophoresis can be used to separate 21-23 nt RNAs from the combination, gel slices comprising the RNA sequences removed and RNAs eluted from the gel slices. Alternatively, non-denaturing methods, such as non-denaturing column chromatography, can be used to isolate the RNA produced. In addition, chromatography (e.g., size exclusion chromatography), glycerol gradient centrifugation, affinity purification with antibody can be used to isolate 21-23 nt RNAs. The RNA-protein complex isolated from the Drosophila in vitro system can also be used directly in the methods described herein (e.g., method of mediating RNAi of mRNA of a gene). Soluble extracts derived from Drosophila embryo that mediate or RNAi are encompassed by the invention. The soluble Drosophila extract can be obtained in a variety of ways. For example, the soluble extract can be obtained from syncytial blastoderm Drosophila embryos as described in Examples 1, 2, and 3. Soluble extracts can be derived from other cells in which RNAi occurs. Alternatively, soluble extracts can be obtained from a cell that does not carry out RNAi. In this instance, the factors needed to mediate RNAi can be introduced into such a cell and the soluble extract is then obtained. The components of the extract can also be chemically synthesized and/or combined using methods known in the art.

Any dsRNA can be used in the methods of the present invention, provided that it has sufficient homology to the targeted gene to mediate RNAi. The sequence of the dsRNA for use in the methods of the present invention need not be known. Alternatively, the dsRNA for use in the present invention can correspond to a known sequence, such as that of an entire gene (one or more) or portion thereof. There is no upper limit on the length of the dsRNA that can be used. For example, the dsRNA can range from about 21 base pairs (bp) of the gene to the full length of the gene or more. In one embodiment, the dsRNA used in the methods of the present invention is about 1000 bp in length. In another embodiment, the dsRNA is about 500 bp in length. In yet another embodiment, the dsRNA is about 22 bp in length.

-17-

The 21 to 23 nt RNAs described herein can be used in a variety of ways. For example, the 21 to 23 nt RNA molecules can be used to mediate RNA interference of mRNA of a gene in a cell or organism. In a specific embodiment, the 21 to 23 nt RNA is introduced into human cells or a human in order to mediate RNA

5   interference in the cells or in cells in the individual, such as to prevent or treat a disease or undesirable condition. In this method, a gene (or genes) that cause or contribute to the disease or undesirable condition is targeted and the corresponding mRNA (the transcriptional product of the targeted gene) is degraded by RNAi. In this embodiment, an RNA of about 21 to about 23 nucleotides that targets the

10   corresponding mRNA (the mRNA of the targeted gene) for degradation is introduced into the cell or organism. The cell or organism is maintained under conditions under which degradation of the corresponding mRNA occurs, thereby mediating RNA interference of the mRNA of the gene in the cell or organism. In a particular embodiment, the method of mediating RNA interference of a gene in a

15   cell comprises combining double-stranded RNA that corresponds to a sequence of the gene with a soluble extract derived from Drosophila embryo, thereby producing a combination. The combination is maintained under conditions in which the double-stranded RNA is processed to RNA of about 21 to about 23 nucleotides. The 21 to 23 nt RNA is then isolated and introduced into the cell or organism. The cell

20   or organism is maintained under conditions in which degradation of mRNA of the gene occurs, thereby mediating RNA interference of the gene in the cell or organism. In the event that the 21-23nt RNA is introduced into a cell in which RNAi, does not normally occur, the factors needed to mediate RNAi are introduced into such a cell or the expression of the needed factors is induced in such a cell. Alternatively, 21 to

25   23 nt RNA produced by other methods (e.g., chemical synthesis, recombinant DNA production) to have a composition the same as or sufficiently similar to a 21 to 23 nt RNA known to mediate RNAi can be similarly used to mediate RNAi. Such 21 to 23 nt RNAs can be altered by addition, deletion, substitution or modification of one or more nucleotides and/or can comprise non-nucleotide materials. A further

30   embodiment of this invention is an ex vivo method of treating cells from an individual to degrade a gene(s) that causes or is associated with a disease or

-18-

undesirable condition, such as leukemia or AIDS. In this embodiment, cells to be treated are obtained from the individual using known methods (e.g., phlebotomy or collection of bone marrow) and 21-23 nt RNAs that mediate degradation of the corresponding mRNA(s) are introduced into the cells, which are then re-introduced

5   into the individual. If necessary, biochemical components needed for RNAi to occur can also be introduced into the cells.

The mRNA of any gene can be targeted for degradation using the methods of mediating interference of mRNA described herein. For example, any cellular or viral mRNA, can be targeted, and, as a result, the encoded protein (e.g., an

10   oncoprotein, a viral protein), expression will be diminished. In addition, the mRNA of any protein associated with/causative of a disease or undesirable condition can be targeted for degradation using the methods described herein.

The present invention also relates to a method of examining the function of a gene in a cell or organism. In one embodiment, an RNA sequence of about 21 to

15   about 23 nucleotides that targets mRNA of the gene for degradation is introduced into the cell or organism. The cell or organism is maintained under conditions under which degradation of mRNA of the gene occurs. The phenotype of the cell or organism is then observed and compared to an appropriate control, thereby providing information about the function of the gene. In another embodiment,

20   double-stranded RNA that corresponds to a sequence of the gene is combined with a soluble extract derived from Drosophila embryo under conditions in which the double-stranded RNA is processed to generate RNA of about 21 to about 23 nucleotides. The RNA of about 21 to about 23 nucleotides is isolated and then introduced into the cell or organism. The cell or organism is maintained under

25   conditions in which degradation of the mRNA of the gene occurs. The phenotype of the cell or organism is then observed and compared to an appropriate control, thereby identifying the function of the gene.

A further aspect of this invention is a method of assessing the ability of 21-23 nt RNAs to mediate RNAi and, particularly, determining which 21-23 nt

30   RNA(s) most efficiently mediate RNAi. In one embodiment of the method, dsRNA corresponding to a sequence of an mRNA to be degraded is combined with

detectably labeled (e.g., end-labeled, such as radiolabeled) mRNA and the soluble extract of this invention, thereby producing a combination. The combination is maintained under conditions under which the double-stranded RNA is processed and the mRNA is degraded. The sites of the most effective cleavage are mapped by

5   comparing the migration of the labeled mRNA cleavage products to markers of known length. 21 mers spanning these sites are then designed and tested for their efficiency in mediating RNAi.

Alternatively, the extract of the present invention can be used to determine whether there is a particular segment or particular segments of the mRNA

10   corresponding to a gene which are more efficiently targeted by RNAi than other regions and, thus, can be especially useful target sites. In one embodiment, dsRNA corresponding to a sequence of a gene to be degraded, labeled mRNA of the gene is combined with a soluble extract that mediates RNAi, thereby producing a combination. The resulting combination is maintained under conditions under

15   which the dsRNA is degraded and the sites on the mRNA that are most efficiently cleaved are identified, using known methods, such as comparison to known size standards on a sequencing gel.

OVERVIEW OF EXAMPLES

Biochemical analysis of RNAi has become possible with the development of

20   the in vitro Drosophila embryo lysate that recapitulates dsRNA-dependent silencing of gene expression described in Example 1 (Tuschl et al., Genes Dev., 13:3191-7 (1999)). In the in vitro system, dsRNA, but not sense or asRNA, targets a corresponding mRNA for degradation, yet does not affect the stability of an unrelated control mRNA. Furthermore, pre-incubation of the dsRNA in the lysate

25   potentiates its activity for target mRNA degradation, suggesting that the dsRNA must be converted to an active form by binding proteins in the extract or by covalent modification (Tuschl et al., Genes Dev., 13:3191-7 (1999)).

The development of a cell-free system from syncytial blastoderm Drosophila embryos that recapitulates many of the features of RNAi is described herein. The

30   interference observed in this reaction is sequence-specific, is promoted by dsRNA,

-20-

but not by single-stranded RNA, functions by specific mRNA degradation, requires
a minimum length of dsRNA and is most efficient with long dsRNA. Furthermore,
preincubation of dsRNA potentiates its activity. These results demonstrate that
RNAi is mediated by sequence specific processes in soluble reactions.

5   As described in Example 2, the in vitro system was used to analyze the
requirements of RNAi and to determine the fate of the dsRNA and the mRNA.
RNAi in vitro requires ATP, but does not require either mRNA translation or
recognition of the 7-methyl-guanosine cap of the targeted mRNA. The dsRNA, but
not single-stranded RNA, is processed in vitro to a population of 21-23 nt species.

10   Deamination of adenosines within the dsRNA does not appear to be required for
formation of the 21-23 nt RNAs. As described herein, the mRNA is cleaved only in
the region corresponding to the sequence of the dsRNA and that the mRNA is
cleaved at 21-23 nt intervals, strongly indicating that the 21-23 nt fragments from
the dsRNA are targeting the cleavage of the mRNA. Furthermore, as described in

15   Examples 3 and 4, when the 21-23 nt fragments are purified and added back to the
soluble extract, they mediate RNA.

          The present invention is illustrated by the following examples, which are not
intended to be limiting in any way.

Example 1   Targeted mRNA degradation by double-stranded RNA in vitro
20   Materials and Methods
RNAs

          Rr-Luc mRNA consisted of the 926 nt Rr luciferase coding sequence flanked
by 25 nt of 5' untranslated sequence from the pSP64 plasmid polylinker and 25 nt of
3' untranslated sequence consisting of 19 nt of pSP64 plasmid polylinker sequence
25   followed by a 6 nt Sac I site. Pp-Luc mRNA contained the 1653 nt Pp luciferase
coding sequence with a Kpn I site introduced immediately before the Pp luciferase
stop codon. The Pp coding sequence was flanked by 5' untranslated sequences
consisting of 21 nt of pSP64 plasmid polylinker followed by the 512 nt of the 5'
untranslated region (UTR) from the Drosophila hunchback mRNA and 3'
30   untranslated sequences consisting of the 562 nt hunchback 3' UTR followed by a 6

nt Sac I site. The hunchback 3' UTR sequences used contained six G-to-U mutations that disrupt function of the Nanos Response Elements in vivo and in vitro. Both reporter mRNAs terminated in a 25 nt poly(A) tail encoded in the transcribed plasmid. For both Rr-Luc and Pp -Luc mRNAs, the transcripts were generated by

5   run-off transcription from plasmid templates cleaved at an Nsi I site that immediately followed the 25 nt encoded poly(A) tail. To ensure that the transcripts ended with a poly(A) tail, the Nsi I-cleaved transcription templates were resected with T4 DNA Polymerase in the presence of dNTPs. The SP6 mMessage mMachine kit (Ambion) was used for in vitro transcription. Using this kit, about 80% of the

10   resulting transcripts are 7-methyl guanosine capped. ^{32}P-radiolabeling was accomplished by including α-^{32}P-UTP in the transcription reaction.

For Pp -Luc, ss, as, and dsRNA corresponded to positions 93 to 597 relative to the start of translation, yielding a 505 bp dsRNA. For Rr -Luc, ss, as, and dsRNA corresponded to positions 118 to 618 relative to the start of translation, yielding a

15   501 bp dsRNA. The Drosophila nanos competitor dsRNA corresponded to positions 122 to 629 relative to the start of translation, yielding a 508 bp dsRNA. ssRNA, asRNA, and dsRNA (diagrammed in Figure 1) were transcribed in vitro with T7 RNA polymerase from templates generated by the polymerase chain reaction. After gel purification of the T7 RNA transcripts, residual DNA template was removed by

20   treatment with RQ1 DNase (Promega). The RNA was then extracted with phenol and chloroform, and then precipitated and dissolved in water.
RNA annealing and native gel electrophoresis.

ssRNA and asRNA (0.5 μM) in 10 mM Tris-HCl (pH 7.5) with 20 mM NaCl were heated to 95 ° C for 1 min then cooled and annealed at room temperature for 12

25   to 16 h. The RNAs were precipitated and resuspended in lysis buffer (below). To monitor annealing, RNAs were electrophoresed in a 2% agarose gel in TBE buffer and stained with ethidium bromide (Sambrook et al., Molecular Cloning. Cold Spring Harbor Laboratory Press, Plainview, NY. (1989)).

-22-

Lysate preparation

Zero- to two-hour old embryos from Oregon R flies were collected on yeasted molasses agar at 25°C. Embryos were dechorionated for 4 to 5 min in 50% (v/v) bleach, washed with water, blotted dry, and transferred to a chilled

5   Potter-Elvehjem tissue grinder (Kontes). Embryos were lysed at 4°C in one ml of lysis buffer (100 mM potassium acetate, 30 mM HEPES-KOH, pH 7.4, 2 mM magnesium acetate) containing 5 mM dithiothreitol (DTT) and 1 mg/ml Pefabloc SC (Boehringer-Mannheim) per gram of damp embryos. The lysate was centrifuged for 25 min at 14,500 x g at 4°C, and the supernatant flash frozen in aliquots in liquid

10 nitrogen and stored at -80°C.

Reaction conditions

Lysate preparation and reaction conditions were derived from those described by Hussain and Leibowitz (Hussain and Leibowitz, Gene 46:13-23 (1986)). Reactions contained 50% (v/v) lysate, mRNAs (10 to 50 pM final

15 concentration), and 10% (v/v) lysis buffer containing the ssRNA, asRNA, or dsRNA (10 nM final concentration). Each reaction also contained 10 mM creatine phosphate, 10 μg/ml creatine phosphokinase, 100 μM GTP, 100 μM UTP, 100 μM CTP, 500 μM ATP, 5 μM DTT, 0.1 U/mL RNasin (Promega), and 100 μM of each amino acid. The final concentration of potassium acetate was adjusted to 100 mM.

20 For standard conditions, the reactions were assembled on ice and then pre-incubated at 25° C for 10 min before adding mRNA. After adding mRNAs, the incubation was continued for an additional 60 min. The 10 min preincubation step was omitted for the experiments in Figures 3A-3C and 5A-5C. Reactions were quenched with four volumes of 1.25x Passive Lysis Buffer (Promega). Pp and Rr luciferase activity

25 was detected in a Monolight 2010 Luminometer (Analytical Luminescence Laboratory) using the Dual-Luciferase Reporter Assay System (Promega).

RNA stability

Reactions with ^{32}P-radiolabeled mRNA were quenched by the addition of 40 volumes of 2x PK buffer (200 mM Tris-HCl, pH 7.5, 25 mM EDTA, 300 mM NaCl,

-23-

2% w/v sodium dodecyl sulfate). Proteinase K (E.M. Merck; dissolved in water) was added to a final concentration of 465 μg/ml. The reactions were then incubated for 15 min at 65° C, extracted with phenol/chloroform/isoamyl alcohol (25:24:1), and precipitated with an equal volume of isopropanol. Reactions were analyzed by electrophoresis in a formaldehyde/agarose (0.8% w/v) gel (Sambrook et al., Molecular Cloning. Cold Spring Harbor Laboratory Press, Plainview, NY. (1989)). Radioactivity was detected by exposing the agarose gel [dried under vacuum onto Nytran Plus membrane (Amersham)] to an image plate (Fujix) and quantified using a Fujix Bas 2000 and Image Gauge 3.0 (Fujix) software.

Commercial lysates

Untreated rabbit reticulocyte lysate (Ambion) and wheat germ extract (Ambion) reactions were assembled according to the manufacturer's directions. dsRNA was incubated in the lysate at 27°C (wheat germ) or 30°C (reticulocyte lysate) for 10 min prior to the addition of mRNAs.

Results and Discussion

To evaluate if dsRNA could specifically block gene expression in vitro, reporter mRNAs derived from two different luciferase genes that are unrelated both in sequence and in luciferin substrate specificity were used: Renilla reniformis (sea pansy) luciferase (Rr-Luc) and Photuris pennsylvanica (firefly) luciferase (Pp-Luc). dsRNA generated from one gene was used to target that luciferase mRNA whereas the other luciferase mRNA was an internal control co-translated in the same reaction. dsRNAs of approximately 500 bp were prepared by transcription of polymerase-chain reaction products from the Rr-Luc and Pp-Luc genes. Each dsRNA began ~100 bp downstream of the start of translation (Figure 1). Sense (ss) and anti-sense (as) RNA were transcribed in vitro and annealed to each other to produce the dsRNA. Native gel electrophoresis of the individual Rr 501 and Pp 505 nt as RNA and ssRNA used to form the Rr and Pp dsRNAs was preformed. The ssRNA, asRNA, and dsRNAs were each tested for their ability to block specifically

-24-

expression of their cognate mRNA but not the expression of the unrelated internal control mRNA.

The ssRNA, asRNA, or dsRNA was incubated for 10 min in a reaction containing Drosophila embryo lysate, then both Pp-Luc and Rr-Luc mRNAs were

5   added and the incubation continued for an additional 60 min. The Drosophila embryo lysate efficiently translates exogenously transcribed mRNA under the conditions used. The amounts of Pp-Luc and Rr-Luc enzyme activities were measured and were used to calculate ratios of either Pp-Luc/Rr-Luc (Figure 2A) or Rr-Luc/Pp-Luc (Figure 2B). To facilitate comparison of different experiments, the

10   ratios from each experiment were normalized to the ratio observed for a control in which buffer was added to the reaction in place of ssRNA, asRNA, or dsRNA.

Figure 2A shows that a 10 nM concentration of the 505 bp dsRNA identical to a portion of the sequence of the Pp-Luc gene specifically inhibited expression of the Pp- Luc mRNA but did not affect expression of the Rr-Luc internal control.

15   Neither ssRNA nor asRNA affected expression of Pp-Luc or the Rr-Luc internal control. Thus, Pp-Luc expression was specifically inhibited by its cognate dsRNA. Conversely, a 10 nM concentration of the 501 bp dsRNA directed against the Rr-Luc mRNA specifically inhibited Rr-Luc expression but not that of the Pp-Luc internal control (Figure 2B). Again, comparable levels of ssRNA or asRNA had little or no

20   effect on expression of either reporter mRNA. On average, dsRNA reduced specific luciferase expression by 70% in these experiments, in which luciferase activity was measured after 1 h incubation. In other experiments in which the translational capacity of the reaction was replenished by the addition of fresh lysate and reaction components, a further reduction in targeted luciferase activity relative to the internal

25   control was observed.

The ability of dsRNA but not asRNA to inhibit gene expression in these lysates is not merely a consequence of the greater stability of the dsRNA (half-life about 2 h) relative to the single-stranded RNAs (half-life ~ 10 min). ssRNA and asRNA transcribed with a 7-methyl guanosine cap were as stable in the lysate as

30   uncapped dsRNA, but do not inhibit gene expression. In contrast, dsRNA formed

-25-

from the capped ssRNA and asRNA specifically blocks expression of the targeted mRNA.

Effective RNAi in Drosophila requires the injection of about 0.2 fmol of dsRNA into a syncytial blastoderm embryo (Kennerdell and Carthew, Cell

5 95:1017-1026 (1998); Carthew, www1.pitt.edu/~carthew/manual/RNAi_Protocol.html (1999)). Since the average volume of a Drosophila embryo is approximately 7.3 nl, this corresponds to an intracellular concentration of about 25 nM (Mazur et al., Cryobiology 25:543-544 (1988)). Gene expression in the Drosophila lysate was inhibited by a comparable

10 concentration of dsRNA (10 nM), but lowering the dsRNA concentration ten-fold decreased the amount of specific interference. Ten nanomolar dsRNA corresponds to a 200-fold excess of dsRNA over target mRNA added to the lysate. To test if this excess of dsRNA might reflect a time- and/or concentration-dependent step in which the input dsRNA was converted to a form active for gene-specific interference, the

15 effect of preincubation of the dsRNA on its ability to inhibit expression of its cognate mRNA was examined. Because the translational capacity of the lysates is significantly reduced after 30 min of incubation at 25°C (unpublished observations), it was desired to ensure that all factors necessary for RNAi remained active throughout the pre-incubation period. Therefore, every 30 min, a reaction

20 containing dsRNA and lysate was mixed with a fresh reaction containing unincubated lysate (Figure 3A). After six successive serial transfers spanning 3 hours of preincubation, the dsRNA, now diluted 64-fold relative to its original concentration, was incubated with lysate and 50 pM of target mRNA for 60 min. Finally, the Pp-Luc and Rr-Luc enzyme levels were measured. For comparison, the

25 input amount of dsRNA (10 nM) was diluted 32-fold in buffer, and its capacity to generate gene-specific dsRNA interference in the absence of any preincubation step was assessed.

The preincubation of the dsRNA in lysate significantly potentiated its capacity to inhibit specific gene expression. Whereas the dsRNA diluted 32-fold

30 showed no effect, the preincubated dsRNA was, within experimental error, as potent as undiluted dsRNA, despite having undergone a 64-fold dilution. Potentiation of

-26-

the dsRNA by preincubation was observed for dsRNAs targeting both the Pp-Luc mRNA (Figure 3B) and the Rr-Luc mRNA (Figure 3C). Taking into account the 64-fold dilution, the activation conferred by preincubation allowed a 156 pM concentration of dsRNA to inhibit 50 pM target mRNA. Further, dilution of the

5   "activated" dsRNA may be effective but has not been tested. We note that although both dsRNAs tested were activated by the preincubation procedure, each fully retained its specificity to interfere with expression only of the mRNA to which it is homologous. Further study of the reactions may provide a route to identifying the mechanism of dsRNA potentiation.

10   One possible explanation for the observation that preincubation of the dsRNA enhances its capacity to inhibit gene expression in these lysates is that specific factors either modify and/or associate with the dsRNA. Accordingly, the addition of increasing amounts of dsRNA to the reaction might titrate such factors and decrease the amount of gene-specific interference caused by a second dsRNA of

15   unrelated sequence. For both Pp-Luc mRNA and Rr-Luc mRNA, addition of increasing concentrations of the unrelated Drosophila nanos dsRNA to the reaction decreased the amount of gene- specific interference caused by dsRNA targeting the reporter mRNA (Figure 4). None of the tested concentrations of nanos dsRNA affected the levels of translation of the untargeted mRNA, demonstrating that the

20   nanos dsRNA specifically titrated factors involved in gene-specific interference and not components of the translational machinery. The limiting factor(s) was titrated by addition of approximately 1000 nM dsRNA, a 200-fold excess over the 5 nM of dsRNA used to produce specific interference.

Interference in vitro might reflect either a specific inhibition of mRNA

25   translation or the targeted destruction of the specific mRNA. To distinguish these two possibilities, the fates of the Pp-Luc and Rr-Luc mRNAs were examined directly using ^{32}P-radiolabeled substrates. Stability of 10 nM Pp-Luc mRNA or Rr-Luc mRNA incubated in lysate with either buffer or 505 bp Pp-dsRNA (10 nM). Samples were deproteinized after the indicated times and the ^{32}P-radiolabeled

30   mRNAs were then resolved by denaturing gel electrophoresis. In the absence of dsRNA, both the Pp-Luc and Rr-Luc mRNAs were stable in the lysates, with ~ 75%

-27-

of the input mRNA remaining after 3 h of incubation. (About 25% of the input mRNA is rapidly degraded in the reaction and likely represents uncapped mRNA generated by the in vitro transcription process.) In the presence of dsRNA (10 nM, 505 bp) targeting the Pp-Luc mRNA, less than 15% of the Pp-Luc mRNA remained
5   after 3 h (Figure 5A). As expected, the Rr-Luc mRNA remained stable in the presence of the dsRNA targeting Pp-Luc mRNA. Conversely, dsRNA (10 nM, 501 bp) targeting the Rr-Luc mRNA caused the destruction of the Rr-Luc mRNA but had no effect on the stability of Pp-Luc mRNA (Figure 5B). Thus, the dsRNA specifically caused accelerated decay of the mRNA to which it is homologous with
10 no effect on the stability of the unrelated control mRNA. This finding indicates that in vivo, at least in Drosophila, the effect of dsRNA is to directly destabilize the target mRNA, not to change the subcellular localization of the mRNA, for example, by causing it to be specifically retained in the nucleus, resulting in non-specific degradation.
15   These results are consistent with the observation that RNAi leads to reduced cytoplasmic mRNA levels in vivo, as measured by in situ hybridization (Montgomery et al., Proc. Natl. Acad. Sci. USA 95:15502-15507 (1998)) and Northern blotting (Ngo et al., Proc. Natl. Acad. Sci. USA 95:14687-14692 (1998)). Northern blot analyses in trypanosomes and hydra suggest that dsRNA typically
20 decreases mRNA levels by less than 90% (Ngo et al., Proc. Natl. Acad. Sci. USA 95:14687-14692 (1998); Lohmann et al., Dev. Biol. 214:211-214 (1999)). The data presented here show that in vitro mRNA levels are reduced 65 to 85% after three hours incubation, an effect comparable with observations in vivo. They also agree with the finding that RNAi in C. elegans is post- transcriptional (Montgomery et al.,
25 Proc. Natl. Acad. Sci. USA 95:15502-15507 (1998)). The simplest explanation for the specific effects on protein synthesis is that it reflects the accelerated rate of RNA decay. However, the results do not exclude independent but specific effects on translation as well as stability.
        In vivo, RNAi appears to require a minimum length of dsRNA (Ngo et al.,
30 Proc. Natl. Acad. Sci., USA, 95:14687-14692 (1998)). The ability of RNA duplexes of lengths 49 bp, 149 bp, 505 bp, and 997 bp (diagrammed in Figure 1) to target the

-28-

degradation of the Pp-Luc mRNA in vitro was assessed. In good agreement with in vivo observations, the 49 bp dsRNA was ineffective in vitro, while the 149 bp dsRNA enhanced mRNA decay only slightly, and both the 505 and 997 bp dsRNAs caused robust mRNA degradation (Figure 5C). 50bp dsRNA targeting other

5   portions of the mRNA cause detectable mRNA degradation, though not as robust as that seen for 500bp dsRNA. Thus, although some short dsRNA do not mediate RNAi, others of approximately the same length, but different composition, will be able to do so.

        Whether the gene-specific interference observed in Drosophila lysates was a

10   general property of cell-free translation systems was examined. The effects of dsRNAs on expression of Pp-Luc and Rr-Luc mRNA were examined in commercially available wheat germ extracts and rabbit reticulocyte lysates. There was no effect of addition of 10 nM of either ssRNA, asRNA, or dsRNA on the expression of either mRNA reporter in wheat germ extracts. In contrast, the addition

15   of 10 nM of dsRNA to the rabbit reticulocyte lysate caused a profound and rapid, non-specific decrease in mRNA stability. For example, addition of Rr-Luc dsRNA caused degradation of both Rr-Luc and Pp-Luc mRNAs within 15 min. The same non-specific effect was observed upon addition of Pp-Luc dsRNA. The non-specific destruction of mRNA induced by the addition of dsRNA to the rabbit reticulocyte

20   lysate presumably reflects the previously observed activation of RNase L by dsRNA (Clemens and Williams, Cell 13:565-572 (1978); Williams et al., Nucleic Acids Res. 6:1335-1350 (1979); Zhou et al., Cell 72:753-765 (1993); Matthews, Interactions between Viruses and the Cellular Machinery for Protein Synthesis. In Translational Control (eds. J. Hershey, M. Mathews and N. Sonenberg), pp. 505-548. Cold

25   Spring Harbor Laboratory Press, Plainview, NY. (1996)). Mouse cell lines lacking dsRNA-induced anti-viral pathways have recently been described (Zhou et al., Virology 258:435-440 (1999)) and may be useful in the search for mammalian RNAi. Although RNAi is known to exist in some mammalian cells (Wianny and Zernicka-Goetz Nat. Cell Biol. 2: 70-75 (2000)), in many mammalian cell types its

30   presence is likely obscured by the rapid induction by dsRNA of non-specific anti-viral responses.

-29-

dsRNA-targeted destruction of specific mRNA is characteristic of RNAi, which has been observed in vivo in many organisms, including Drosophila. The system described above recapitulates in a reaction in vitro many aspects of RNAi. The targeted mRNA is specifically degraded whereas unrelated control mRNAs

5   present in the same solution are not affected. The process is most efficient with dsRNAs greater than 150 bp in length. The dsRNA-specific degradation reaction in vitro is probably general to many, if not all, mRNAs since it was observed using two unrelated genes.

The magnitude of the effects on mRNA stability in vitro described herein are

10   comparable with those reported in vivo (Ngo et al., Proc. Natl. Acad. Sci., USA, 95:14687-14692 (1998); Lohmann et al., Dev. Biol., 214:211-214 (1999). However, the reaction in vitro requires an excess of dsRNA relative to mRNA. In contrast, a few molecules of dsRNA per cell can inhibit gene expression in vivo (Fire et al., Nature, 391: 806-811 (1998); Kennerdell and Carthew, Cell, 95:1017-1026 (1998)).

15   The difference between the stoichiometry of dsRNA to target mRNA in vivo and in vitro should not be surprising in that most in vitro reactions are less efficient than their corresponding in vivo processes. Interestringly, incubation of the dsRNA in the lysate greatly potentiated its activity for RNAi, indicating that it is either modified or becomes associated with other factors or both. Perhaps a small number of molecules

20   is effective in inhibiting the targeted mRNA in vivo because the injected dsRNA has been activated by a process similar to that reported here for RNAi in Drosophila lysates.

Example 2   Double-Stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals

25   Methods and Material
In vitro RNAi

In vitro RNAi reactions and lysate preparation were as described in Example 1 (Tuschl et al., Genes Dev., 13:3191-7 (1999)) except that the reaction contained 0.03 g/ml creatine kinase, 25 µM creatine phosphate (Fluka), and 1 mM ATP.

-30-

Creatine phosphate was freshly dissolved at 500 mM in water for each experiment. GTP was omitted from the reactions, except in Figures 2 and 3.

RNA Synthesis.

Pp-luc and Rr-luc mRNAs and Pp- and Rr-dsRNAs (including dsRNA 'B' in Figure 6) were synthesized by in vitro transcription as described previously (Tuschl et al., Genes Dev., 13:3191-7 (1999)). To generate transcription templates for dsRNA 'C,' the 5′ sense RNA primer was gcgtaatacgactcactataGAACAAAGGAAACGGATGAT (SEQ ID NO: 2) and the 3′ sense RNA primer was GAAGAAGTTATTCTCCAAAA (SEQ ID NO: 3); the 5′ asRNA primer was gcgtaatacgactcactataGAAGAAGTTATTCTCCAAAA (SEQ ID NO: 4)and the 3′ asRNA primer was GAACAAAGGAAACGGATGAT (SEQ ID NO: 5). For dsRNA 'A' the 5′ sense RNA primer was gcgtaatacgactcactataGTAGCGCGGTGTATTATACC (SEQ ID NO: 6)and the 3′ sense RNA primer was GTACAACGTCAGGTTTACCA (SEQ ID NO: 7); the 5′ asRNA primer was gcgtaatacgactcactataGTACAACGTCAGGTTTACCA (SEQ ID NO: 8)and the 3′ asRNA primer was GTAGCGCGGTGTATTATACC (SEQ ID NO: 9) (lowercase, T7 promoter sequence).

mRNAs were 5′-end-labeled using guanylyl transferase (Gibco/BRL), S-adenosyl methionine (Sigma), and α-^{32}P-GTP (3000 Ci/mmol; New England Nuclear) according to the manufacturer's directions. Radiolabeled RNAs were purified by poly(A) selection using the Poly(A) Tract III kit (Promega). Nonradioactive 7-methyl- guanosine- and adenosine-capped RNAs were synthesized in in vitro transcription reactions with a 5-fold excess of 7-methyl-G(5′)ppp(5′)G or A(5′)ppp(5′)G relative to GTP. Cap analogs were purchased from New England Biolabs.

ATP depletion and Protein Synthesis Inhibition

ATP was depleted by incubating the lysate for 10 minutes at 25°C with 2 mM glucose and 0.1 U/ml hexokinase (Sigma). Protein synthesis inhibitors were purchased from Sigma and dissolved in absolute ethanol as 250-fold concentrated

-31-

stocks. The final concentrations of inhibitors in the reaction were: anisomycin, 53 mg/ml; cycloheximide, 100 mg/ml; chloramphenicol, 100 mg/ml. Relative protein synthesis was determined by measuring the activity of Rr luciferase protein produced by translation of the Rr-luc mRNA in the RNAi reaction after 1 hour as

5   described previously (Tuschl et al., Genes Dev., 13:3191-7 (1999)).

Analysis of dsRNA Processing

Internally α-^{32}P-ATP-labeled dsRNAs (505 bp Pp-luc or 501 Rr-luc) or 7-methyl-guanosine-capped Rr-luc antisense RNA (501 nt) were incubated at 5 nM final concentration in the presence or absence of unlabeled mRNAs in Drosophila lysate

10   for 2 hours in standard conditions. Reactions were stopped by the addition of 2x proteinase K buffer and deproteinized as described previously (Tuschl et al., Genes Dev., 13:3191- 3197 (1999)). Products were analyzed by electrophoresis in 15% or 18% polyacrylamide sequencing gels. Length standards were generated by complete RNase T1 digestion of α-^{32}P-ATP-labeled 501 nt Rr-luc sense RNA and asRNA.

15   For analysis of mRNA cleavage, 5'-^{32}P-radiolabeled mRNA (described above) was incubated with dsRNA as described previously (Tuschl et al., Genes Dev., 13:3191- 3197 (1999)) and analyzed by electrophoresis in 5% (Figure 5B) and 6% (Figure 6C) polyacrylamide sequencing gels. Length standards included commercially available RNA size standards (FMC Bioproducts) radiolabeled with

20   guanylyl transferase as described above and partial base hydrolysis and RNase T1 ladders generated from the 5'-radiolabeled mRNA.

Deamination Assay

Internally α-^{32}P-ATP-labeled dsRNAs (5 nM) were incubated in Drosophila lysate for 2 hours at standard conditions. After deproteinization, samples were run

25   on 12% sequencing gels to separate full-length dsRNAs from the 21-23 nt products. RNAs were eluted from the gel slices in 0.3 M NaCl overnight, ethanol-precipitated, collected by centrifugation, and redissolved in 20 µl water. The RNA was hydrolyzed into nucleoside 5 -phosphates with nuclease P1 (10 µl reaction containing 8 µl RNA in water, 30 mM KOAc pH 5.3, 10 mM ZnSO$_4$, 10 µg or 3

-32-

units nuclease P1, 3 hours, 50° C). Samples (1 ml) were co-spotted with non-radioactive 5 -mononucleotides [0.05 O.D. units (A_{260}) of pA, pC, pG, pI, and pU] on cellulose HPTLC plates (EM Merck) and separated in the first dimension in isobutyric acid/25% ammonia/water (66/1/33, v/v/v) and in the second dimension in

5   0.1M sodium phosphate, pH 6.8/ammonium sulfate/1-propanol (100/60/2, v/w/v; Silberklang et al., 1979). Migration of the non- radioactive internal standards was determined by UV-shadowing.

Results and Discussion
RNAi Requires ATP

10   As described in Example 1, Drosophila embryo lysates faithfully recapitulate RNAi (Tuschl et al., Genes Dev., 13:3191-7 (1999)). Previously, dsRNA-mediated gene silencing was monitored by measuring the synthesis of luciferase protein from the targeted mRNA. Thus, these RNAi reactions contained an ATP-regenerating system, needed for the efficient translation of the mRNA. To test if ATP was, in

15   fact, required for RNAi, the lysates were depleted for ATP by treatment with hexokinase and glucose, which converts ATP to ADP, and RNAi was monitored directly by following the fate of ^{32}P-radiolabeled Renilla reniformis luciferase (Rr-luc) mRNA (Figure 6). Treatment with hexokinase and glucose reduced the endogenous ATP level in the lysate from 250 μM to below 10 μM. ATP

20   regeneration required both exogenous creatine phosphate and creatine kinase, which acts to transfer a high-energy phosphate from creatine phosphate to ADP. When ATP-depleted extracts were supplemented with either creatine phosphate or creatine kinase separately, no RNAi was observed. Therefore, RNAi requires ATP in vitro. When ATP, creatine phosphate, and creatine kinase were all added together to

25   reactions containing the ATP-depleted lysate, dsRNA-dependent degradation of the Rr-luc mRNA was restored (Figure 6). The addition of exogenous ATP was not required for efficient RNAi in the depleted lysate, provided that both creatine phosphate and creatine kinase were present, demonstrating that the endogenous concentration (250 mM) of adenosine nucleotide is sufficient to support RNAi.

30   RNAi with a Photinus pyralis luciferase (Pp-luc) mRNA was also ATP-dependent.

-33-

The stability of the Rr-luc mRNA in the absence of Rr-dsRNA was reduced in ATP-depleted lysates relative to that observed when the energy regenerating system was included, but decay of the mRNA under these conditions did not display the rapid decay kinetics characteristic of RNAi in vitro, nor did it generate the stable

5   mRNA cleavage products characteristic of dsRNA-directed RNAi. These experiments do not establish if the ATP requirement for RNAi is direct, implicating ATP in one or more steps in the RNAi mechanism, or indirect, reflecting a role for ATP in maintaining high concentrations of another nucleoside triphosphate in the lysate.

10   Translation Is Not Required for RNAi In Vitro

The requirement for ATP suggested that RNAi might be coupled to mRNA translation, a highly energy-dependent process. To test this possibility, various inhibitors of protein synthesis were added to the reaction by preparing a denaturing agarose-gel analysis of 5′-32P-radiolabeled Pp-luc mRNA after incubation for

15   indicated times in a standard RNAi reaction with and without protein synthesis inhibitors. The eukaryotic translation inhibitors anisomycin, an inhibitor of initial peptide bond formation, cycloheximide, an inhibitor of peptide chain elongation, and puromycin, a tRNA mimic which causes premature termination of translation (Cundliffe, Antibiotic Inhibitors of Ribosome Function. In The Molecular Basis of

20   Antibiotic Action, E. Gale, E. Cundliffe, P. Reynolds, M. Richmond and M. Warning, eds. (New York: Wiley), pp. 402-547. (1981)) were tested. Each of these inhibitors reduced protein synthesis in the Drosophila lysate by more than 1,900-fold (Figure 7A). In contrast, chloramphenicol, an inhibitor of Drosophila mitochondrial protein synthesis (Page and Orr-Weaver, Dev. Biol., 183:195-207 (1997)), had no

25   effect on translation in the lysates (Figure 7A). Despite the presence of anisomycin, cycloheximide, or chloramphenicol, RNAi proceeded at normal efficiency. Puromycin also did not perturb efficient RNAi. Thus, protein synthesis is not required for RNAi in vitro.

Translational initiation is an ATP-dependent process that involves

30   recognition of the 7-methyl guanosine cap of the mRNA (Kozak, Gene, 234:187-208

-34-

(1999); Merrick and Hershey, The Pathway and Mechanism of Eukaryotic Protein Synthesis. In Translational Control, J. Hershey, M. Mathews and N. Sonenberg, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 31-69 (1996)). The Drosophila lysate used to support RNAi in vitro also recapitulates the

5   cap-dependence of translation; Pp-luc mRNA with a 7-methyl-guanosine cap was translated greater than ten-fold more efficiently than was the same mRNA with an A(5′)ppp(5′)G cap (Figure 7B). Both RNAs were equally stable in the Drosophila lysate, showing that this difference in efficiency cannot be merely explained by more rapid decay of the mRNA with an adenosine cap (see also Gebauer et al., EMBO J.,

10   18:6146-54 (1999)). Although the translational machinery can discriminate between Pp-luc mRNAs with 7- methyl-guanosine and adenosine caps, the two mRNAs were equally susceptible to RNAi in the presence of Pp-dsRNA (Figure 7C). These results suggest that steps in cap recognition are not involved in RNAi.

dsRNA Is Processed to 21-23 nt Species

15   RNAs 25 nt in length are generated from both the sense and anti-sense strands of genes undergoing post-transcriptional gene silencing in plants (Hamilton and Baulcombe, Science, 286:950-2 (1999)). Denaturing acrylamide-gel analysis of the products formed in a two-hour incubation of uniformly ^{32}P-radiolabeled dsRNAs and capped asRNA in lysate under standard RNAi conditions, in the presence or

20   absence of target mRNAs. It was found that dsRNA is also processed to small RNA fragments. When incubated in lysate, approximately 15% of the input radioactivity of both the 501 bp Rr-dsRNA and the 505 bp Pp-dsRNA appeared in 21 to 23 nt RNA fragments. Because the dsRNAs are more than 500 bp in length, the 15% yield of fragments implies that multiple 21-23 nt RNAs are produced from each

25   full-length dsRNA molecule. No other stable products were detected. The small RNA species were produced from dsRNAs in which both strands were uniformly ^{32}P-radiolabeled. Formation of the 21-23 nt RNAs from the dsRNA did not require the presence of the corresponding mRNA, demonstrating that the small RNA species is generated by processing of the dsRNA, rather than as a product of

-35-

dsRNA-targeted mRNA degradation. It was noted that 22 nucleotides corresponds to two turns of an A-form RNA-RNA helix.

When dsRNAs radiolabeled within either the sense or the anti-sense strand were incubated with lysate in a standard RNAi reaction, 21-23 nt RNAs were generated with comparable efficiency. These data support the idea that the 21-23 nt RNAs are generated by symmetric processing of the dsRNA. A variety of data support the idea that the 21-23 nt RNA is efficiently generated only from dsRNA and is not the consequence of an interaction between single-stranded RNA and the dsRNA. First, a ^{32}P-radiolabeled 505 nt Pp-luc sense RNA or asRNA was not efficiently converted to the 21-23 nt product when it was incubated with 5 nM nonradioactive 505 bp Pp- dsRNA. Second, in the absence of mRNA, a 501 nt 7-methyl-guanosine-capped Rr- asRNA produced only a barely detectable amount of 21-23 nt RNA (capped single- stranded RNAs are as stable in the lysate as dsRNA, Tuschl et al., Genes Dev., 13:3191- 7 (1999)), probably due to a small amount of dsRNA contaminating the anti-sense preparation. However, when Rr-luc mRNA was included in the reaction with the ^{32}P- radiolabeled, capped Rr-asRNA, a small amount of 21-23 nt product was generated, corresponding to 4% of the amount of 21-23 nt RNA produced from an equimolar amount of Rr-dsRNA. This result is unlikely to reflect the presence of contaminating dsRNA in the Rr-asRNA preparation, since significantly more product was generated from the asRNA in the presence of the Rr-luc mRNA than in the absence. Instead, the data suggest that asRNA can interact with the complementary mRNA sequences to form dsRNA in the reaction and that the resulting dsRNA is subsequently processed to the small RNA species. Rr-asRNA can support a low level of bona fide RNAi in vitro (see below), consistent with this explanation.

It was next asked if production of the 21-23 nt RNAs from dsRNA required ATP. When the 505 bp Pp-dsRNA was incubated in a lysate depleted for ATP by treatment with hexokinase and glucose, 21-23 nt RNA was produced, albeit 6 times slower than when ATP was regenerated in the depleted lysate by the inclusion of creatine kinase and creatine phosphate. Therefore, ATP may not be required for production of the 21-23 nt RNA species, but may instead simply enhance its

-36-

formation. Alternatively, ATP may be required for processing of the dsRNA, but at a concentration less than that remaining after hexokinase treatment. The molecular basis for the slower mobility of the small RNA fragments generated in the ATP-depleted lysate is not understood.

5   Wagner and Sun (Wagner and Sun, Nature, 391:744-745 (1998)) and Sharp (Sharp, Genes Dev., 13:139-41 (1999)) have speculated that the requirement for dsRNA in gene silencing by RNAi reflects the involvement of a dsRNA-specific adenosine deaminase in the process. dsRNA adenosine deaminases unwind dsRNA by converting adenosine to inosine, which does not base-pair with uracil. dsRNA

10   adenosine deaminases function in the post-transcriptional editing of mRNA (for review see Bass, Trends Biochem. Sci., 22:157-62 (1997)). To test for the involvement of dsRNA adenosine deaminase in RNAi, the degree of conversion of adenosine to inosine in the 501 bp Rr-luc and 505 bp Pp-luc dsRNAs after incubation with Drosophila embryo lysate in a standard in vitro RNAi reaction was

15   examined. Adenosine deamination in full-length dsRNA and the 21-23 nt RNA species was assessed by two-dimensional thin-layer chromatography. Inorganic phosphate (P_i,) was produced by the degradation of mononucleotides by phosphatases that contaminate commercially available nuclease P1 (Auxilien et al., J. Mol. Biol., 262:437-458 (1996)). The degree of adenosine deamination in the

20   21-23 nt species was also determined. The full-length dsRNA radiolabeled with [^{32}P]-adenosine was incubated in the lysate, and both the full-length dsRNA and the 21-23 nt RNA products were purified from a denaturing acrylamide gel, cleaved to mononucleotides with nuclease P1, and analyzed by two-dimensional thin- layer chromatography.

25   A significant fraction of the adenosines in the full-length dsRNA were converted to inosine after 2 hours (3.1% and 5.6% conversion for Pp-luc and Rr-luc dsRNAs, respectively). In contrast, only 0.4% (Pp-dsRNA) or 0.7% (Rr-dsRNA) of the adenosines in the 21-23 nt species were deaminated. These data imply that fewer than 1 in 27 molecules of the 21-23 nt RNA species contain an inosine.

30   Therefore, it is unlikely that dsRNA-dependent adenosine deamination within the 21-23 nt species is required for its production.

-37-

asRNA Generates a Small Amount of RNAi in vitro

When mRNA was ^{32}P-radiolabeled within the 5′-7-methyl-guanosine cap, stable 5′ decay products accumulated during the RNAi reaction. Such stable 5′decay products were observed for both the Pp-luc and Rr-luc mRNAs when they were incubated with their cognate dsRNAs. Previously, it was reported that efficient RNAi does not occur when asRNA is used in place of dsRNA (Tuschl et al., Genes Dev., 13:3191-7 (1999)). Nevertheless, mRNA was measurably less stable when incubated with asRNA than with buffer (Figures 8A and 8B). This was particularly evident for the Rr-luc mRNA: approximately 90% of the RNA remained intact after a 3-hour incubation in lysate, but only 50% when asRNA was added. Less than 5% remained when dsRNA was added. Interestingly, the decrease in mRNA stability caused by asRNA was accompanied by the formation of a small amount of the stable 5′-decay products characteristic of the RNAi reaction with dsRNA. This finding parallels the observation that a small amount of 21- 23 nt product formed from the asRNA when it was incubated with the mRNA (see above) and lends strength to the idea that asRNA can enter the RNAi pathway, albeit inefficiently.

mRNA Cleavage Sites Are Determined by the Sequence of the dsRNA

The sites of mRNA cleavage were examined using three different dsRNAs, 'A,' 'B,' and 'C,' displaced along the Rr-luc sequence by approximately 100 nts. Denaturing acrylamide-gel analysis of the stable, 5′-cleavage products produced after incubation of the Rr-luc mRNA for the indicated times with each of the three dsRNAs, 'A,' 'B,' and 'C,' or with buffer (Ø) was performed. The positions of these relative to the Rr-luc mRNA sequence are shown in Figure 9. Each of the three dsRNAs was incubated in a standard RNAi reaction with Rr-luc mRNA ^{32}P-radiolabeled within the 5′-cap. In the absence of dsRNA, no stable 5′-cleavage products were detected for the mRNA, even after 3 hours of incubation in lysate. In contrast, after a 20-minute incubation, each of the three dsRNAs produced a ladder of bands corresponding to a set of mRNA cleavage products characteristic for that particular dsRNA. For each dsRNA, the stable, 5′ mRNA cleavage products were restricted to the region of the Rr-luc mRNA that corresponded to the dsRNA

-38-

(Figures 9 and 10). For dsRNA 'A,' the lengths of the 5'- cleavage products ranged from 236 to just under ~750 nt; dsRNA 'A' spans nucleotides 233 to 729 of the Rr-luc mRNA. Incubation of the mRNA with dsRNA 'B' produced mRNA 5'-cleavage products ranging in length from 150 to ~600 nt; dsRNA 'B' spans

5   nucleotides 143 to 644 of the mRNA. Finally, dsRNA 'C' produced mRNA cleavage products from 66 to ~500 nt in length. This dsRNA spans nucleotides 50 to 569 of the Rr-luc mRNA. Therefore, the dsRNA not only provides specificity for the RNAi reaction, selecting which mRNA from the total cellular mRNA pool will be degraded, but also determines the precise positions of cleavage along the mRNA

10   sequence.

The mRNA Is Cleaved at 21-23 Nucleotide Intervals

To gain further insight into the mechanism of RNAi, the positions of several mRNA cleavage sites for each of the three dsRNAs were mapped (Figure 10). High resolution denaturing acrylamide-gel analysis of a subset of the 5'-cleavage products

15   described above was performed. Remarkably, most of the cleavages occurred at 21-23 nt intervals (Figure 10). This spacing is especially striking in light of our observation that the dsRNA is processed to a 21-23 nt RNA species and the finding of Hamilton and Baulcombe that a 25 nt RNA correlates with post-transcriptional gene silencing in plants (Hamilton and Baulcombe, Science, 286:950-2 (1999)). Of

20   the 16 cleavage sites we mapped (2 for dsRNA 'A,' 5 for dsRNA 'B,' and 9 for dsRNA 'C'), all but two reflect the 21-23 nt interval. One of the two exceptional cleavages was a weak cleavage site produced by dsRNA 'C' (indicated by an open blue circle in Figure 10). This cleavage occurred 32 nt 5' to the next cleavage site. The other exception is particularly intriguing. After four cleavages spaced 21-23 nt

25   apart, dsRNA 'C' caused cleavage of the mRNA just nine nt 3' to the previous cleavage site (red arrowhead in Figure 10). This cleavage occurred in a run of seven uracil residues and appears to "reset" the ruler for cleavage; the next cleavage site was 21-23 nt 3' to the exceptional site. The three subsequent cleavage sites that we mapped were also spaced 21-23 nt apart. Curiously, of the sixteen cleavage sites

30   caused by the three different dsRNAs, fourteen occur at uracil residues. The

-39-

significance of this finding is not understood, but it suggests that mRNA cleavage is determined by a process which measures 21-23 nt intervals and which has a sequence preference for cleavage at uracil. Results show that the 21-23 nt RNA species produced by incubation of ~500 bp dsRNA in the lysate caused sequence-

5   specific interference in vitro when isolated from an acrylamide gel and added to a new RNAi reaction in place of the full-length dsRNA.

A Model for dsRNA-directed mRNA Cleavage

Without wishing to be bound by theory, the biochemical data described herein, together with recent genetic experiments in C. elegans and Neurospora

10   (Cogoni and Macino, Nature, 399:166-9 (1999); Grishok et al., Science, 287: 2494-7 (2000); Ketting et al., Cell, 99:133-41 (1999); Tabara et al., Cell, 99:123-32 (1999)), suggest a model for how dsRNA targets mRNA for destruction (Figure 11). In this model, the dsRNA is first cleaved to 21-23 nt long fragments in a process likely to involve genes such as the C. elegans loci rde-1 and rde-4. The resulting fragments,

15   probably as short asRNAs bound by RNAi-specific proteins, would then pair with the mRNA and recruit a nuclease that cleaves the mRNA. Alternatively, strand exchange could occur in a protein-RNA complex that transiently holds a 21-23 nt dsRNA fragment close to the mRNA. Separation of the two strands of the dsRNA following fragmentation might be assisted by an ATP-dependent RNA helicase,

20   explaining the observed ATP enhancement of 21-23 nt RNA production.

It is likely that each small RNA fragment produces one, or at most two, cleavages in the mRNA, perhaps at the 5′ or 3′ ends of the 21-23 nt fragment. The small RNAs may be amplified by an RNA-directed RNA polymerase such as that encoded by the ego-1 gene in C. elegans (Smardon et al., Current Biology,

25   10:169-178 (2000)) or the qde-1 gene in Neurospora (Cogoni and Macino, Nature, 399:166-9 (1999)), producing long-lasting post-transcriptional gene silencing in the absence of the dsRNA that initiated the RNAi effect. Heritable RNAi in C. elegans requires the rde-1 and rde-4 genes to initiate, but not to persist in subsequent generations. The rde-2, rde- 3, and mut-7 genes in C. elegans are required in the

30   tissue where RNAi occurs, but are not required for initiation of heritable RNAi

-40-

(Grishok et al., Science, in press 2000). These 'effector' genes (Grishok et al., Science, in press 2000) are likely to encode proteins functioning in the actual selection of mRNA targets and in their subsequent cleavage. ATP may be required at any of a number of steps during RNAi, including complex formation on the

5   dsRNA, strand dissociation during or after dsRNA cleavage, pairing of the 21-23 nt RNAs with the target mRNA, mRNA cleavage, and recycling of the targeting complex. Testing these ideas with the in vitro RNAi system will be an important challenge for the future. Some genes involved in RNAi are also important for transposon silencing and co-suppresion. Co-suppression is a broad biological

10   phenomenon spanning plants, insects and perhaps humans. The most likely mechanism in Drosophila melanogaster is transcriptional silencing (Pal-Bhanra et al, Cell 99: 35-36. Thus, 21-23 nt fragments are likely to be involved in transcriptional control, as well as in post-transcriptional cotrol.

Example 3   Isolated 21-23 mers caused sequence-specific interference when
15   added to a new RNAi reaction

Isolation of 21-23 nt fragments from incubation reaction of 500 bp dsRNA in lysate.

Double-stranded RNA (500 bp from) was incubated at 10 nM concentration in Drosophila embryo lysate for 3 h at 25° C under standard conditions as described herein. After deproteinization of the sample, the 21-23 nt reaction products were

20   separated from unprocessed dsRNA by denaturing polyacrylamide (15%) gel electrophoresis. For detection of the non-radiolabeled 21-23 nt fragments, an incubation reaction with radiolabeled dsRNA was loaded in a separate lane of the same gel. Gel slices containing the non-radioactive 21-23 nt fragments were cut out and the 21-23 nt fragments were eluted from the gel slices at 4° C overnight in 0.4

25   ml 0.3 M NaCl. The RNA was recovered from the supernatant by ethanol precipitation and centrifugation. The RNA pellet was dissolved in 10 μl of lysis buffer. As control, gel slices slightly above and below the 21-23 nt band were also cut out and subjected to the same elution and precipitation procedures. Also, a non-incubated dsRNA loaded on the 15% gel and a gel slice corresponding to 21-23

30   nt fragments was cut out and eluted. All pellets from the control experiments were

-41-

dissolved in 10 µl lysis buffer. The losses of RNA during recovery from gel slices by elution are approx. 50%.

Incubation of purified 21-23 nt fragments in a translation-based RNAi assay

1 µl of the eluted 21-23 mer or control RNA solution was used for a standard

5   10 µl RNAi incubation reaction (see above). The 21-23 mers were preincubated in the lysate containing reaction mixture for 10 or 30 min before the addition of the target and control mRNA. During pre-incubation, proteins involved in RNA interference may re- associate with the 21-23 mers due to a specific signal present on these RNAs. The incubation was continued for another hour to allow translation of

10   the target and control mRNAs. The reaction was quenched by the addition of passive lysis buffer (Promega), and luciferase activity was measured. The RNA interference is the expressed as the ratio of target to control luciferase activity normalized by an RNA-free buffer control. Specific suppression of the target gene was observed with either 10 or 30 minutes pre- incubation. The suppression was reproducible and

15   reduced the relative ratio of target to control by 2-3 fold. None of the RNA fragments isolated as controls showed specific interference. For comparison, incubation of 5 nM 500 bp dsRNA (10 min pre- incubation) affects the relative ratio of control to target gene approx. 30-fold.

Stability of isolated 21-23 nt fragments in a new lysate incubation reaction.

20   Consistent with the observation of RNAi mediated by purified 21-23 nt RNA fragment, it was found that 35% of the input 21-23 nt RNA persists for more than 3 h in such an incubation reaction. This suggests that cellular factors associate with the deproteinized 21-23 nt fragments and reconstitute a functional mRNA-degrading particle. Signals connected with these 21-23 nt fragments, or their possible double

25   stranded nature or specific lengths are likely responsible for this observation. The 21-23 nt fragments have a terminal 3' hydroxyl group, as evidenced by altered mobility on a sequencing gel following periodate treatment and beta-elimination.

-42-

Example 4   21-23-mers purified by non-denaturing methods caused
sequence-specific interference when added to a new RNAi reaction.

Fifty nanomolar double-stranded RNA (501 bp Rr-luc dsRNA, as described
in example 1) was incubated in a 1 ml in vitro reaction with lysate at 25°C (see
5   example 1). The reaction was then stopped by the addition of an equal volume of 2x
PK buffer (see example 1) and proteinase K was added to a final concentration of
1.8 μg/μl. The reaction was incubated for an additional 1 h at 25°C, phenol
extracted, and then the RNAs were precipitated with 3 volumes of ethanol. The
ethanol precipitate was collected by centrifugation, and the pellet was resuspended
10   in 100 μl of lysis buffer and applied to a Superdex HR 200 10/30 gel filtration
column (Pharmacia) run in lysis buffer at 0.75 ml/min. 200 μl fractions were
collected from the column. Twenty μl of 3 M sodium acetate and 20 μg glycogen
was added to each fraction, and the RNA was recovered by precipitation with 3
volumes of ethanol. The precipitates were resuspended in 30 μl of lysis buffer.
15   Column profiles following the fractionation of 32P-labeled input RNA are shown in
Figure 13A.

One microliter of each resuspended fraction was tested in a 10 μl standard in
vitro RNAi reaction (see example 1). This procedure yields a concentration of RNA
in the in vitro RNAi reaction that is approximately equal to the concentration of that
20   RNA species in the original reaction prior to loading on the column. The fractions
were preincubated in the lysate containing reaction mixture for 30 min before the
addition of 10 nM Rr-luc mRNA target and 10 nM Pp-luc control mRNA. During
pre-incubation, proteins involved in RNA interference may re-associate with the
21-23-mers due to a specific signal present on these RNAs. The incubation was
25   continued for another three hours to allow translation of the target and control
mRNAs. The reaction was quenched by the addition of passive lysis buffer
(Promega), and luciferase activity was measured. The suppression of Rr-luc mRNA
target expression by the purified 21-23 nt fragments was reproducible and reduced
the relative ratio of target to control by >30-fold, an amount comparable to a 50 nM
30   500 bp dsRNA control. Suppression of target mRNA expression was specific: little
or no effect on the expression of the Pp-luc mRNA control was observed.

-43-

The data show that the both the fractions containing uncleaved dsRNA (fractions 3 - 5) or long, partially cleaved dsRNA (fractions 7 - 13) and the fractions containing the fully processed 21-23 nt siRNAs (fractions 41 - 50) mediate effective RNA interference in vitro (Figure 13B). Suppression of target mRNA expression was specific: little or no effect on the expression of the Pp-luc mRNA control was observed (Figure 13C). These data, together with those in the earlier examples, demonstrate that the 21-23 nt siRNAs are (1) true intermediates in the RNAi pathway and (2) effective mediators of RNA interference in vitro.

Example 5　　　21-nucleotide siRNA duplexes mediate RNA interference in human tissue cultures

Methods

RNA preparation

21 nt RNAs were chemically synthesized using Expedite RNA phosphoramidites and thymidine phosphoramidite (Proligo, Germany). Synthetic oligonucleotides were deprotected and gel-purified (Elbashir, S.M., Lendeckel, W. & Tuschl, T., *Genes & Dev. 15*, 188-200 (2001)), followed by Sep-Pak C18 cartridge (Waters, Milford, MA, USA) purification (Tuschl, t., *et al., Biochemistry, 32:*11658-11668 (1993)). The siRNA sequences targeting GL2 (Acc. X65324) and GL3 luciferase (Acc. U47296) corresponded to the coding regions 153-173 relative to the first nucleotide of the start codon, siRNAs targeting RL (Acc. AF025846) corresponded to region 119-129 after the start codon. Longer RNAs were transcribed with T7 RNA polymerase from PCR products, followed by gel and Sep-Pak purification. The 49 and 484 bp GL2 or GL3 dsRNAs corresponded to position 113-161 and 113-596, respectively, relative to the start of translation; the 50 and 501 bp RL dsRNAs corresponded to position 118-167 and 118-618, respectively. PCR templates for dsRNA synthesis targeting humanized GFP (hG) were amplified from pAD3 (Kehlenbach, R.H., *et al., J. Cell Biol., 141:*863-874 (1998)), whereby 50 and 501 bp hG dsRNA corresponded to position 118-167 and 118-618, respectively, to the start codon.

-44-

For annealing of siRNAs, 20 µM single strands were incubated in annealing buffer (100 mM potassium acetate, 30 mM HEPES-KOH at pH 7.4, 2 mM magnesium acetate) for 1 min at 90°C followed by 1 h at 37 °C. The 37 °C incubation step was extended overnight for the 50 and 500 bp dsRNAs, and these annealing reactions were performed at 8.4 µM and 0.84 µM strand concentrations, respectively.

Cell culture

S2 cells were propagated in Schneider's *Drosophila* medium (Life Technologies) supplemented with 10% FBS, 100 units/ml penicillin, and 100 µg/ml streptomycin at 25 °C. 293, NIH/3T3, HeLa S3, COS-7 cells were grown at 37 °C in Dulbecco's modified Eagle's medium supplemented with 10% FBS, 100 units/ml penicillin, and 100 µg/ml streptomycin. Cells were regularly passaged to maintain exponential growth. 24 h before transfection at approx. 80% confluency, mammalian cells were trypsinized and diluted 1:5 with fresh medium without antibiotics (1-3 × 10^5 cells/ml) and transferred to 24-well plates (500 µl/well). S2 cells were not trypsinized before splitting. Transfection was carried out with Lipofectamine 2000 reagent (Life Technologies) as described by the manufacturer for adherent cell lines. Per well, 1.0 µg pGL2-Control (Promega) or pGL3-Control (Promega), 0.1 µg pRL-TK (Promega), and 0.28 µg siRNA duplex or dsRNA, formulated into liposomes, were applied; the final volume was 600 µl per well. Cells were incubated 20 h after transfection and appeared healthy thereafter. Luciferase expression was subsequently monitored with the Dual luciferase assay (Promega). Transfection efficiencies were determined by fluorescence microscopy for mammalian cell lines after co-transfection of 1.1 µg hGFP-encoding pAD3[22] and 0.28 µg invGL2 siRNA, and were 70-90%. Reporter plasmids were amplified in XL-1 Blue (Strategene) and purified using the Qiagen EndoFree Maxi Plasmid Kit.

Results

RNA interference (RNAi) is the process of sequence-specific, post-transcriptional gene silencing in animals and plants, initiated by double-stranded

-45-

RNA (dsRNA) homologous in sequence to the silenced gene (Fire, A., *Trends Genet., 15:*358-363 (1999); Sharp, P.A. & Zamore, P.D., *Science, 287:*2431-2433 (2000); Sijen, T. & Kooter, J.M., *Bioessays, 22:*520-531 (2000); Bass, B.L., *Cell, 101:*235-238 (2000); Hammond, S.M., *et al., Nat. Rev. Genet., 2:*110-119 (2001)).

5   The mediators of sequence-specific mRNA degradation are 21 and 22 nt small interfering RNAs (siRNAs) generated by RNase III cleavage from longer dsRNAs[6-10] (Hamilton, A.J. &Baulcombe, D.C., *Science, 286:*950-952 (1999); Hammond, S.M., *et al., Nature, 404:*293-296 (2000); Zamore, P.D., *et al., Cell, 101:*25-33 (2000); Bernstein, E., *et al., Naature, 409:*363-366 (2001); Elbashir, S.M., *et al., Genes &*

10   *Dev., 15:*188-200 (2001)). As shown herein, 21 nt siRNA duplexes are able to specifically suppress reporter gene expression in multiple mammalian tissue cultures, including human embryonic kidney (293) and HeLa cells. In contrast to 50 or 500 bp dsRNAs, siRNAs do not activate the interferon response. These results indicate that siRNA duplexes are a general tool for sequence-specific inactivation of

15   gene function in mammalian cells.

        Base-paired 21 and 22 nt siRNAs with overhanging 3' ends mediate efficient sequence-specific mRNA degradation in lysates prepared from *D. melanogaster* embryos (Elbashir, S.M., *et al., Genes & Dev., 15:*188-200 (2001)). To test whether siRNAs are also capable of mediating RNAi in tissue culture, 21 nt siRNA duplexes

20   with symmetric 2 nt 3' overhangs directed against reporter genes coding for sea pansy (*Renilla reniformis*) and two sequence variants of firefly (*Photinus pyralis,* GL2 and GL3) luciferases (Figures 14A, 14B) were constructed. The siRNA duplexes were co-transfected with the reporter plasmid combinations pGL2/pRL or pGL3/pRL, into *D. melanogaster Schneider S2* cells or mammalian cells using

25   cationic liposomes. Luciferase activities were determined 20 h after transfection. In all cell lines tested, specific reduction of the expression of the reporter genes in the presence of cognate siRNA duplexes was observed (Figures 15A-15J). Remarkably, the absolute luciferase expression levels were unaffected by non-cognate siRNAs, indicating the absence of harmful side effects by 21 nt RNA duplexes (e.g. Figures

30   16A-16D, for HeLa cells). In *D. melanogaster* S2 cells (Figures 15A, 15B), the specific inhibition of luciferases was complete, and similar to results previously

-46-

obtained for longer dsRNAs (Hammond, S.M., et al., Nature, 404:293-296 (2000);
Caplen, N.J., et al., Gene, 252:95-105 (2000); Clemens, M & Williams, B., Cell,
13:565-572 (1978); Ui-Tei, K., et al., FEBS Letters, 479:79-82 (2000)). In
mammalian cells, where the reporter genes were 50- to 100-fold stronger expressed,

5   the specific suppression was less complete (Figures 15C-15J). GL2 expression was
reduced 3- to 12-fold, GL3 expression 9- to 25-fold, and RL expression 1- to 3-fold,
in response to the cognate siRNAs. For 293 cells, targeting of RL luciferase by RL
siRNAs was ineffective, although GL2 and GL3 targets responded specifically
(Figures 15I, 15J). It is likely that the lack of reduction of RL expression in 293 cells

10   is due to its 5- to 20-fold higher expression compared to any other mammalian cell
line tested and/or to limited accessibility of the target sequence due to RNA
secondary structure or associated proteins. Nevertheless, specific targeting of GL2
and GL3 luciferase by the cognate siRNA duplexes indicated that RNAi is also
functioning in 293 cells.

15   The 2 nt 3' overhang in all siRNA duplexes, except for uGL2, was composed
of (2'-deoxy) thymidine. Substitution of uridine by thymidine in the 3' overhang was
well tolerated in the D. melanogaster in vitro system, and the sequence of the
overhang was uncritical for target recognition (Elbashir, S.M., et al., Genes & Dev.,
15:188-200 (2001)). The thymidine overhang was chosen, because it is supposed to

20   enhance nuclease resistance of siRNAs in the tissue culture medium and within   ·
transfected cells. Indeed, the thymidine-modified GL2 siRNA was slightly more
potent than the unmodified uGL2 siRNA in all cell lines tested (Figures 15A, 15C,
15E, 15G, 15I). It is conceivable that further modifications of the 3' overhanging
nucleotides will provide additional benefits to the delivery and stability of siRNA

25   duplexes.

        In co-transfection experiments, 25 nM siRNA duplexes with respect to the
final volume of tissue culture medium were used (Figures 15A-15J, 16A-16F).
Increasing the siRNA concentration to 100 nM did not enhance the specific silencing
effects, but started to affect transfection efficiencies due to competition for liposome

30   encapsulation between plasmid DNA and siRNA. Decreasing the siRNA
concentration to 1.5 nM did not reduce the specific silencing effect, even though the

-47-

siRNAs were now only 2- to 20-fold more concentrated than the DNA plasmids. This indicates that siRNAs are extraordinarily powerful reagents for mediating gene silencing, and that siRNAs are effective at concentrations that are several orders of magnitude below the concentrations applied in conventional antisense or ribozyme gene targeting experiments.

5

In order to monitor the effect of longer dsRNAs on mammalian cells, 50 and 500 bp dsRNAs cognate to the reporter genes were prepared. As non-specific control, dsRNAs from humanized GFP (hG) (Kehlenbach, R.H., *et al., J. Cell Biol., 141*:863-874 (1998)) was used. When dsRNAs were co-transfected, in identical amounts (not concentrations) to the siRNA duplexes, the reporter gene expression was strongly and unspecifically reduced. This effect is illustrated for HeLa cells as a representative example (Figures 16A-16D). The absolute luciferase activities were decreased unspecifically 10- to 20-fold by 50 bp dsRNA, and 20- to 200-fold by 500 bp dsRNA co-transfection, respectively. Similar unspecific effects were observed for COS-7 and NIH/3T3 cells. For 293 cells, a 10- to 20-fold unspecific reduction was observed only for 500 bp dsRNAs. Unspecific reduction in reporter gene expression by dsRNA > 30 bp was expected as part of the interferon response (Matthews, M., Interactions between viruses and the cellular machinery for protein synthesis in *Translational Control* (eds., Hershey, J., Matthews,M. & Sonenberg, N.) 505-548 (Cold Spring Harbor Laboratory Press, Plainview, NY; 1996); Kumar, M. & Carmichael, G.G., *Microbiol. Mol. Biol. Rev., 62*:1415-1434 (1998); Stark, G.R., *et al., Annu. Rev. Biochem., 67*:227-264 (1998)). Surprisingly, despite the strong unspecific decrease in reporter gene expression, additional sequence-specific, dsRNA-mediated silencing were reproducibly detected. The specific silencing effects, however, were only apparent when the relative reporter gene activities were normalized to the hG dsRNA controls (Figures 16E, 16F). A 2- to 10-fold specific reduction in response to cognate dsRNA was observed, also in the other three mammalian cell lines tested. Specific silencing effects with dsRNAs (356-1662 bp) were previously reported in CHO-K1 cells, but the amounts of dsRNA required to detect a 2- to 4-fold specific reduction were about 20-fold higher than in our experiments (Ui-Tei, K., *et al., FEBS Letters, 479*:79-82 (2000)). Also, CHO-K1

cells appear to be deficient in the interferon response. In another report, 293, NIH/3T3, and BHK-21 cells were tested for RNAi using luciferase/lacZ reporter combinations and 829 bp specific lacZ or 717 bp unspecific GFP dsRNA (Caplen, N.J., *et al., Gene, 252:*95-105 (2000)). The failure of detecting RNAi in this case is

5 likely due to the less sensitive luciferase/lacZ reporter assay and the length differences of target and control dsRNA. Taken together, the results described herein indicate that RNAi is active in mammalian cells, but that the silencing effect is difficult to detect if the interferon system is activated by dsRNA >30 bp.

The mechanism of the 21 nt siRNA-mediated interference process in

10 mammalian cells remains to be uncovered, and silencing may occur post-transcriptional and/or transcriptional. In *D. melanogaster* lysate, siRNA duplexes mediate post-transcriptional gene silencing by reconstitution of a siRNA-protein complexes (siRNPs), which are guiding mRNA recognition and targeted cleavage (Hammond, S.M., *et al., Nature, 404:*293-296 (2000); Zamore, P.D., *et al., Cell,*

15 *101:*25-33 (2000); Elbashir, S.M., *et al., Genes & Dev., 15:*188-200 (2001)). In plants, dsRNA-mediated post-transcriptional silencing has also been linked to RNA-directed DNA methylation, which may also be directed by 21 nt siRNAs (Wassenegger, M., *Plant Mol. Biol, 43:*203-220 (2000); Finnegan, E.J., *et al., Curr. Biol., 11:*R99-R102 (2000)). Methylation of promoter regions can lead to

20 transcriptional silencing (Metter, M.F., *et al., EMBO J., 19:*5194-5201 (2000)), but methylation in coding sequences must not (Wang, M.-B., *RNA, 7:*16-28 (2001)). DNA methylation and transcriptional silencing in mammals are well-documented processes (Kass, S.U., *et al., Trends Genet., 13:*444-449 (1997); Razin, A., *EMBO J, 17:*4905-4908 (1998)), yet they have not been linked to post-transcriptional

25 silencing. Methylation in mammals is predominantly directed towards CpG residues. Because there is no CpG in the RL siRNA, but RL siRNA mediates specific silencing in mammalian tissue culture, it is unlikely that DNA methylation is critical for our observed silencing process. In summary, described herein, is siRNA-mediated gene silencing in mammalian cells. The use of 21 nt siRNAs holds great

30 promise for inactivation of gene function in human tissue culture and the development of gene-specific therapeutics.

-49-

While this invention has been particularly shown and described with reference to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims

-50-

CLAIMS

What is claimed is:

1.   Isolated RNA of from about 21 to about 23 nucleotides that mediates RNA
        interference of an mRNA to which it corresponds.

5   2.   Isolated RNA of claim 1 that comprises a terminal 3' hydroxyl group.

3.   Isolated RNA of claim 1 which is chemically synthesized RNA or an analog
        of a naturally occurring RNA.

4.   An analog of isolated RNA of claim 1, wherein the analog differs from the
        RNA of claim 1 by the addition, deletion, substitution or alteration of one or
10   more nucleotides.

5.   Isolated RNA of from about 21 to about 23 nucleotides that inactivates a
        corresponding gene by transcriptional silencing.

6.   A soluble extract that mediates RNA interference.

7.   The soluble extract of Claim 6, wherein the extract is derived from
15   Drosophila embryos.

8.   The soluble extract of Claim 7 wherein the extract is derived from syncytial
        blastoderm Drosophila embryos.

9.   A method of producing RNA of from about 21 to about 23 nucleotides in
        length comprising:
20   (a)   combining double-stranded RNA with a soluble extract that mediates
                RNA interference, thereby producing a combination; and

-51-

    (b)   maintaining the combination of a) under conditions in which the double-stranded RNA is processed to RNA of from about 21 to about 23 nucleotides in length.

10.   The method of Claim 9, wherein the soluble extract is derived from syncytial blastoderm Drosophila embryos.

11.   The method of Claim 9 further comprising isolating the RNA of from about 21 to about 23 nucleotides from the combination.

12.   RNA of about 21 to about 23 nucleotides produced by the method of Claim 9.

13.   A method of producing RNA of from about 21 to about 23 nucleotides in length that mediates RNA interference of mRNA of a gene to be degraded, comprising:

    (a)   combining double-stranded RNA that corresponds to a sequence of the gene to be degraded with a soluble extract that mediates RNA interference, thereby producing a combination; and

    (b)   maintaining the combination of (a) under conditions under which the double-stranded RNA is processed to RNA of from about 21 to about 23 nucleotides that mediates RNA interference of the mRNA of the gene to be degraded, thereby producing RNA of from about 21 to about 23 nucleotides that mediates RNA interference of the mRNA.

14.   The method of Claim 13, wherein the soluble extract is derived from syncytial blastoderm Drosophila embryos.

15.   The method of Claim 13 further comprising isolating RNA of from about 21 to about 23 nucleotides from the combination.

-52-

16. Isolated RNA of from about 21 to about 23 nucleotides produced by the method of Claim 15.

17. A method of mediating RNA interference of mRNA of a gene in a cell or organism comprising:

   (a) introducing RNA of from about 21 to about 23 nucleotides which targets the mRNA of the gene for degradation into the cell or organism;

   (b) maintaining the cell or organism produced in (a) under conditions under which degradation of the mRNA occurs, thereby mediating RNA interference of the mRNA of the gene in the cell or organism.

18. The method of Claim 17 wherein the RNA of (a) is a chemically synthesized RNA or an analog of naturally occurring RNA.

19. The method of Claim 17, wherein the gene encodes a cellular mRNA or a viral mRNA.

20. A method of mediating RNA interference of mRNA of a gene in a cell or organism in which RNA interference occurs, comprising:

   (a) combining double-stranded RNA that corresponds to a sequence of the gene with a soluble extract that mediates RNA interference, thereby producing a combination;

   (b) maintaining the combination produced in (a) under conditions under which the double- stranded RNA is processed to RNA of from about 21 to about 23 nucleotides, thereby producing RNA of from about 21 to about 23 nucleotides;

   (c) isolating RNA of from about 21 to about 23 nucleotides produced in (b);

   (d) introducing RNA isolated in (c) into the cell or organism; and

-53-

(e)   maintaining the cell or organism produced in (d) under conditions
under which degradation of mRNA of the gene occurs, thereby
mediating RNA interference of the mRNA of the gene in the cell or
organism.

5   21.   The method of Claim 20, wherein the soluble extract is derived from
syncytial blastoderm Drosophila embryos.

22.   The method of Claim 20, wherein the RNA is isolated using gel
electrophoresis.

23.   A method of mediating RNA interference of mRNA of a gene in a cell or
10   organism in which RNA interference occurs, comprising: (a) introducing into
the cell or organism RNA of from about 21 to about 23 nucleotides that
mediates RNA interference of mRNA of the gene, thereby producing a cell
or organism that contains the RNA and (b) maintaining the cell or organism
that contains the RNA under conditions under which RNA interference
15   occurs, thereby mediating RNA interference of mRNA of the gene in the cell
or organism.

24.   The method of claim 23, wherein the RNA of from about 21 to about 23
nucleotides is chemically synthesized RNA or an analog of RNA that
mediates RNA interference.

20   25.   The method of Claim 23, wherein the gene encodes a cellular mRNA or a
viral mRNA.

26.   A knockdown cell or organism generated by the method of claim 23.

27.   The knockdown cell or organism of claim 26, wherein the cell or organism
mimics a disease.

-54-

28.   A method of examining the function of a gene in a cell or organism comprising:

(a)   introducing RNA of from about 21 to about 23 nucleotides that targets mRNA of the gene for degradation into the cell or organism, thereby producing a test cell or test organism;

(b)   maintaining the test cell or test organism under conditions under which degradation of mRNA of the gene occurs, thereby producing a test cell or test organism in which mRNA of the gene is degraded; and

(c)   observing the phenotype of the test cell or test organism produced in (b) and, optionally, comparing the phenotype observed to that of an appropriate control cell or control organism, thereby providing information about the function of the gene.

29.   The method of Claim 28 wherein the RNA introduced in (a) is chemically synthesized or an analog of RNA that mediates RNA interference.

30.   A method of examining the function of a gene in a cell or organism comprising

(a)   combining double-stranded RNA that corresponds to a sequence of the gene with a soluble extract that mediates RNA interference, thereby producing a combination;

(b)   maintaining the combination produced in (a) under conditions under which the double- stranded RNA is processed to RNA of about 21 to about 23 nucleotides, whereby RNA of about 21 to about 23 nucleotides is produced;

(c)   isolating RNA of about 21 to about 23 nucleotides produced in (b);

(d)   introducing the RNA isolated in (c) into the cell or organism, thereby producing a test cell or test organism;

-55-

(e)   maintaining the test cell or test organism under conditions under which degradation of mRNA of the gene occurs, thereby producing a test cell or test organism in which mRNA of the gene is degraded; and

5   (f)   observing the phenotype of the test cell or test organism produced in (e) and, optionally, comparing the phenotype observed to that of an appropriate control, thereby providing information about the function of the gene.

31.   The method of claim 30, wherein the RNA comprises a terminal 3' hydroxyl
10   group.

32.   The method of claim 30, wherein the soluble extract is derived from syncytial blastoderm Drosophila embryos.

33.   The method of claim 30, wherein the RNA is isolated using gel electrophoresis.

15 34.   A composition comprising biochemical components of a Drosophila cell that process dsRNA to RNA of about 21 to about 23 nucleotides and a suitable carrier.

35.   A composition comprising biochemical components of a cell that target mRNA of a gene to be degraded by RNA of about 21 to about 23
20   nucleotides.

36.   A method of treating a disease or condition associated with the presence of a protein in an individual comprising administering to the individual RNA of from about 21 to about 23 nucleotides that targets the mRNA of the protein for degradation.

-56-

37.   The method of claim 36 wherein RNA of from about 21 to about 23 nucleotides is chemically synthesized or an analog of RNA that mediates RNA interference.

38.   A method of assessing whether an agent acts on a gene product comprising:

5   (a)   introducing RNA of from about 21 to about 23 nucleotides which targets the mRNA of the gene for degradation into a cell or organism;

         (b)   maintaining the cell or organism of (a) under conditions in which degradation of the mRNA occurs,

         (c)   introducing the agent into the cell or organism of (b); and

10   (d)   determining whether the agent has an effect on the cell or organism, wherein if the agent has no effect on the cell or organism then the agent acts on the gene product or on a biological pathway that involves the gene product.

39.   The method of claim 38, wherein the RNA of from about 21 to about 23

15   nucleotides is chemically synthesized or an analog of RNA that mediates RNA interference.

40.   A method of assessing whether a gene product is a suitable target for drug discovery comprising:

         (a)   introducing RNA of from about 21 to about 23 nucleotides which

20   targets the mRNA of the gene for degradation into a cell or organism;

         (b)   maintaining the cell or organism of (a) under conditions in which degradation of the mRNA occurs resulting in decreased expression of the gene; and

         (c)   determining the effect of the decreased expression of the gene on the

25   cell or organism, wherein if decreased expression has an effect, then the gene product is a target for drug discovery.

41.   The method of claim 40, wherein the RNA of from about 21 to about 23 nucleotides is synthetic RNA or an analog of RNA that mediates RNA interference.

42.   A gene identified by the sequencing of endogenous 21 to 23 nucleotide RNA molecules that mediate RNA interference.

43.   A pharmaceutical composition comprising RNA of from about 21 to about 23 nucleotides that mediates RNA interference and an appropriate carrier.

44.   A method of producing knockdown cells, comprising introducing into cells in which a gene is to be knocked down RNA of about 21 to about 23 nt that targets the mRNA corresponding to the gene and maintaining the resulting cells under conditions under which RNAi occurs, resulting in degradation of the mRNA of the gene, thereby producing knockdown cells.

45.   The method of claim 44, wherein the RNA of about 21 to about 23 nucleotides is synthetic RNA or an analog of RNA that mediates RNA interference.

46.   A method of identifying target sites within mRNA that are efficiently cleaved by the RNAi process, comprising combining dsRNA corresponding to a sequence of a gene to be degraded, labeled mRNA corresponding to the gene and a soluble extract that mediates RNA interference, thereby producing a combination; maintaining the combination under conditions under which the dsRNA is degraded and identifying sites in the mRNA that are efficiently cleaved.

47.   A method of identifying 21-23 nt RNAs that efficiently mediate RNAi, wherein said 21-23 nt RNAs span the target sites identified within the mRNA by the method of claim 46.

48.   RNA of claim 16, isolated using gel electrophoresis.

49.   RNA of claim 16, isolated using non-denaturing methods.

50.   RNA of claim 16, isolated using non-denaturing column chromatography.
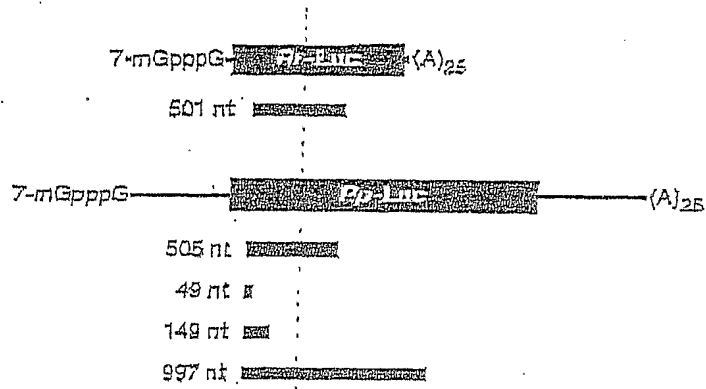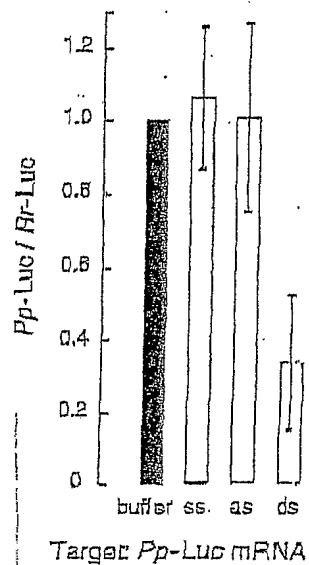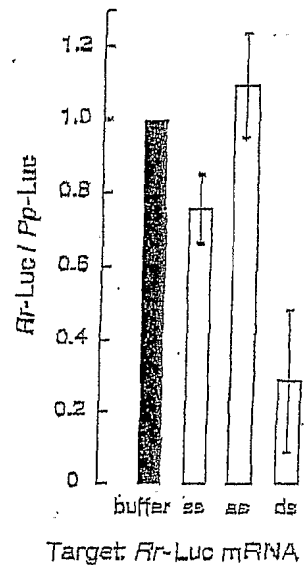
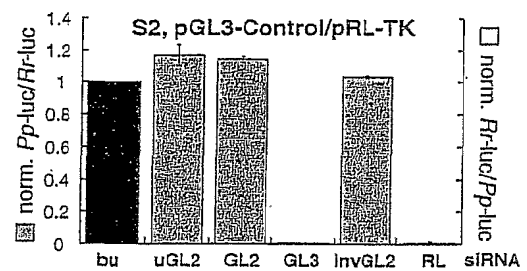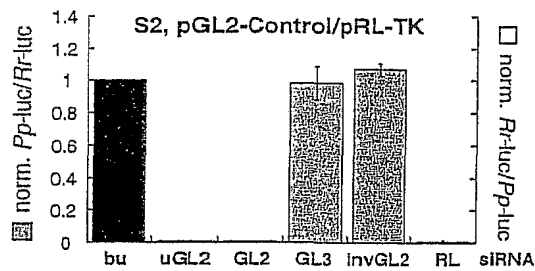Figure 1

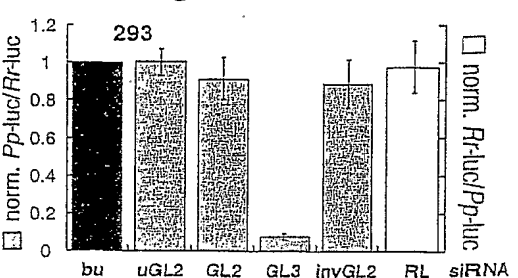

Figure 2A

Figure 2B
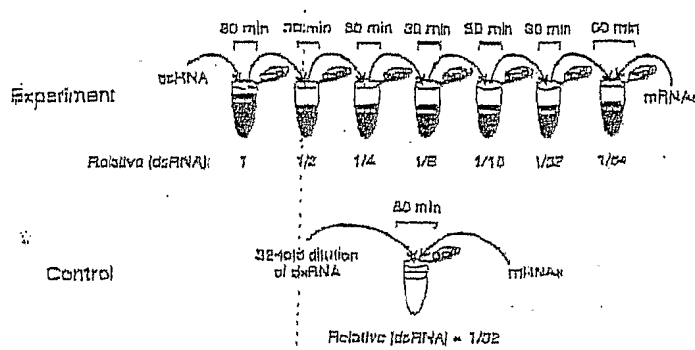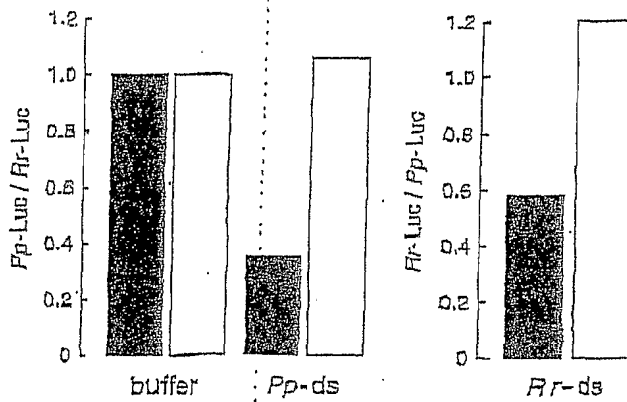
Figure 3A

Figure 3B

Figure 3C
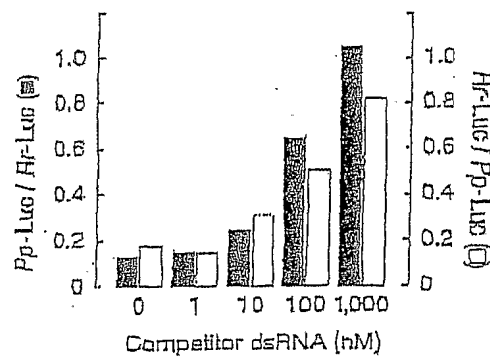
Figure 4
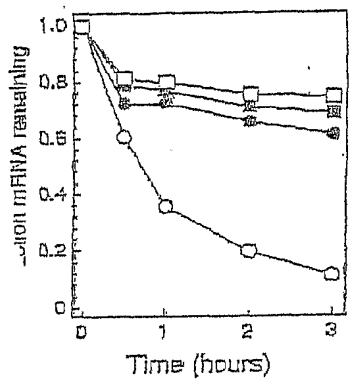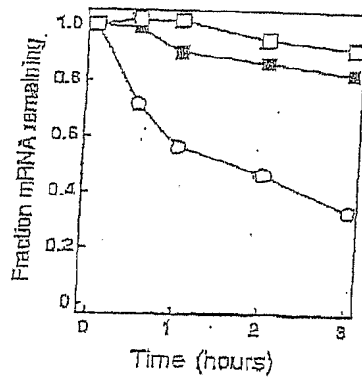
Figure. 5A　　　　　　　　Figure 5B　　　　　　　　Figure 5C
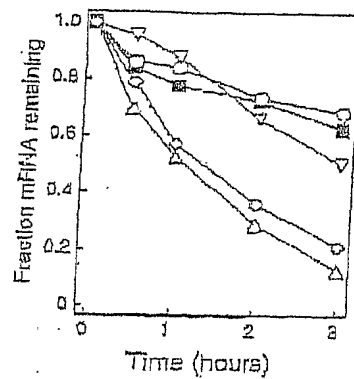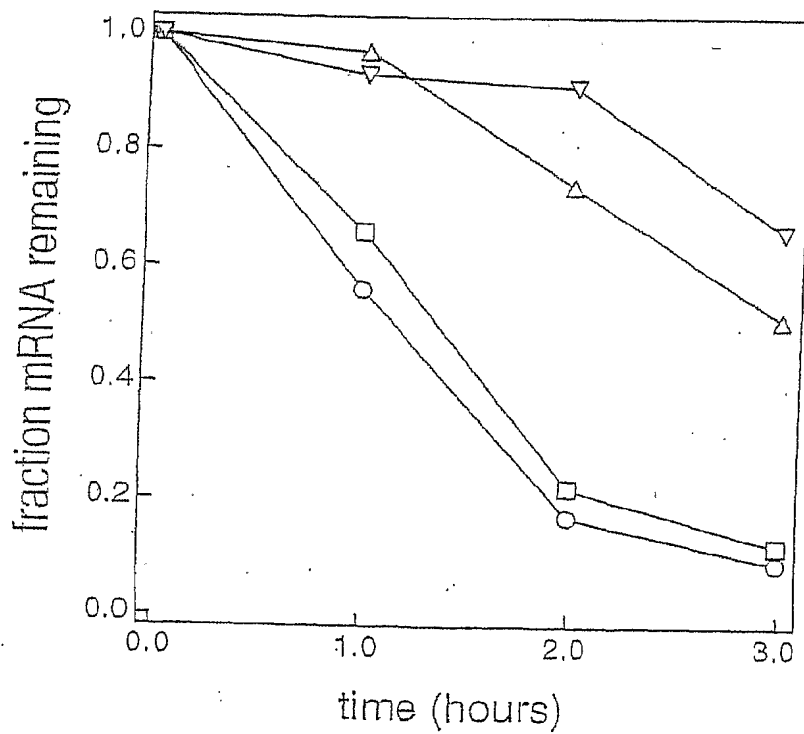
Figure 6
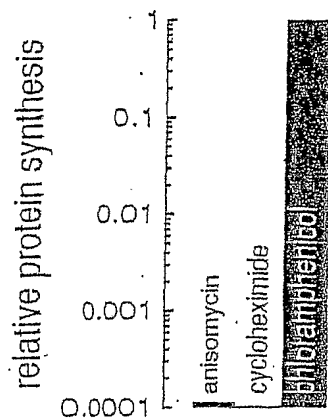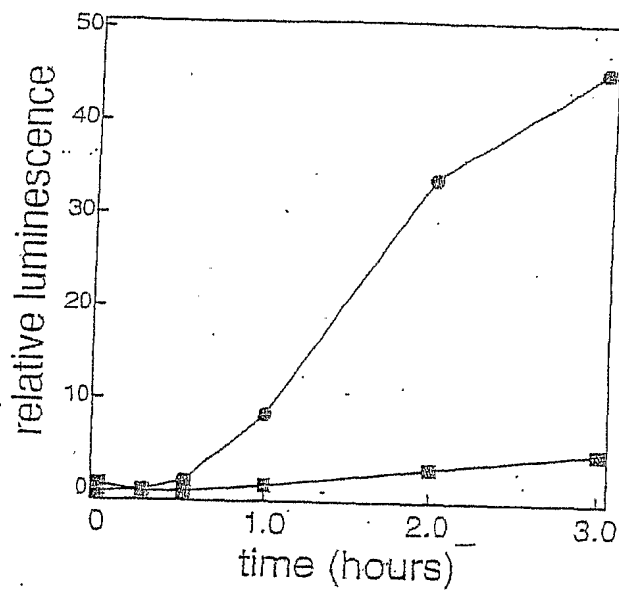
Figure 7A

Figure 7B
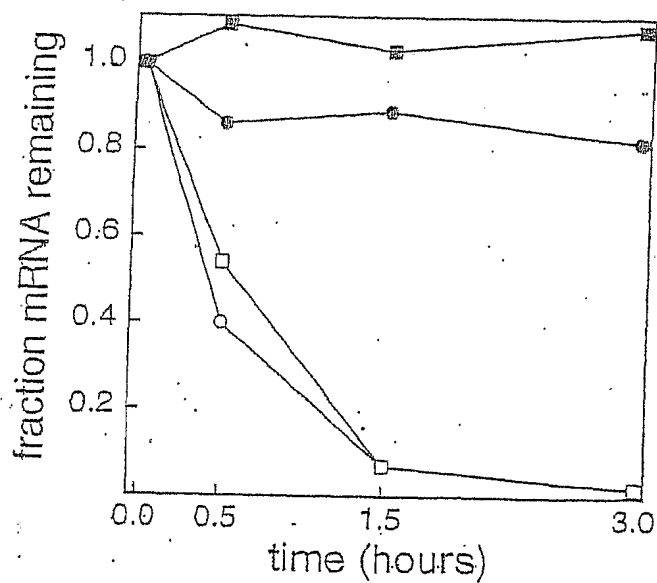
Figure 7C
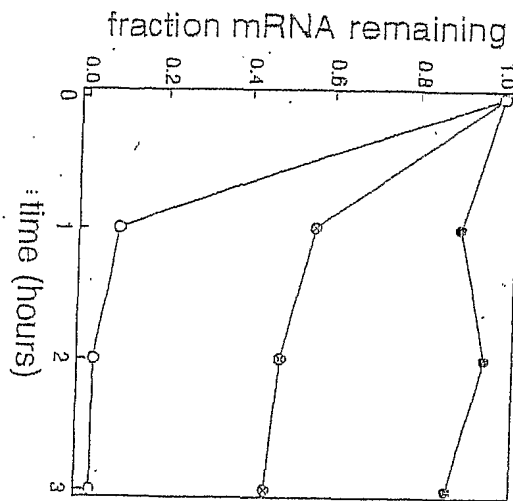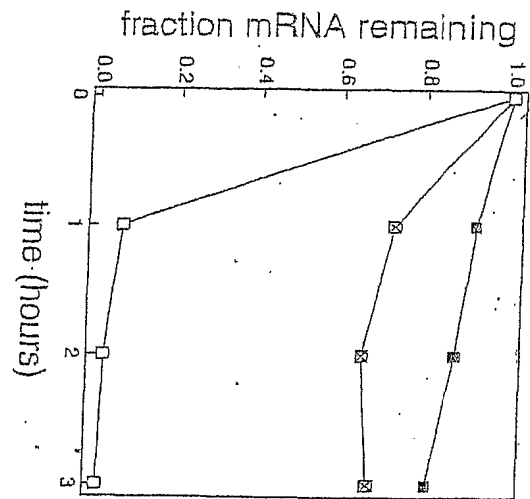
Figure 8A

Figure 8B
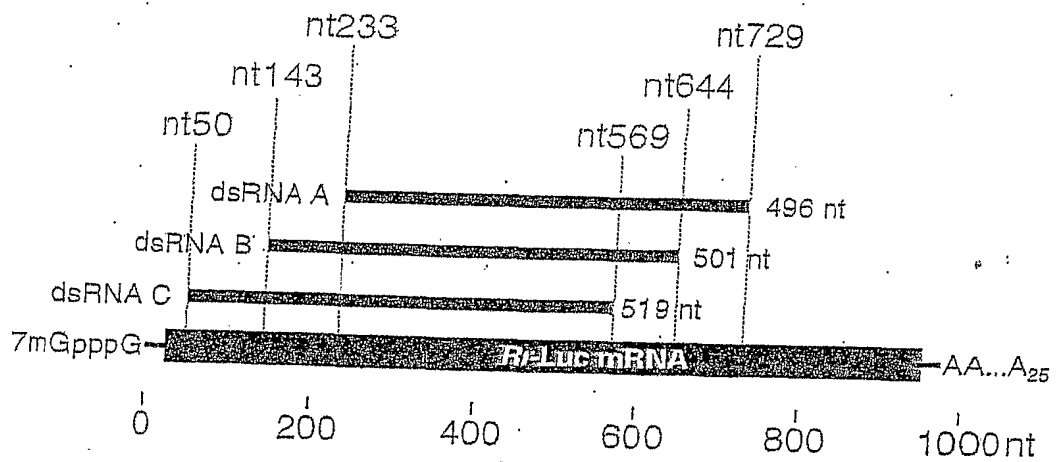
Figure 9

Figure 10

7mG*pppGAAUACAAGCUUGGGCCUAGCCACCAUGACUUCGAAAGUUUAUGAUCC

AGAACAAAGGAAACGGAUGAUAACUGGUCCGCAGUGGUGGGCCAGAUG

UAAACAAAUGAAUGUUCUUGAUUCAUUUAUUAAUUAUUAUGAUUCAGAAA

AACAUGCAGAAAAUGCUGUUAUUUUUUUACAUGGUAACGCGGCCUCUU

CUUAUUUAUGGCGACAUGUUGUGCCACAUAUUGAGCCAGUAGCGCGGU

GUAUUAUACCAGACCUUAUUGGUAU...

Sequence-specific gene silencing by 21-23 nt fragments
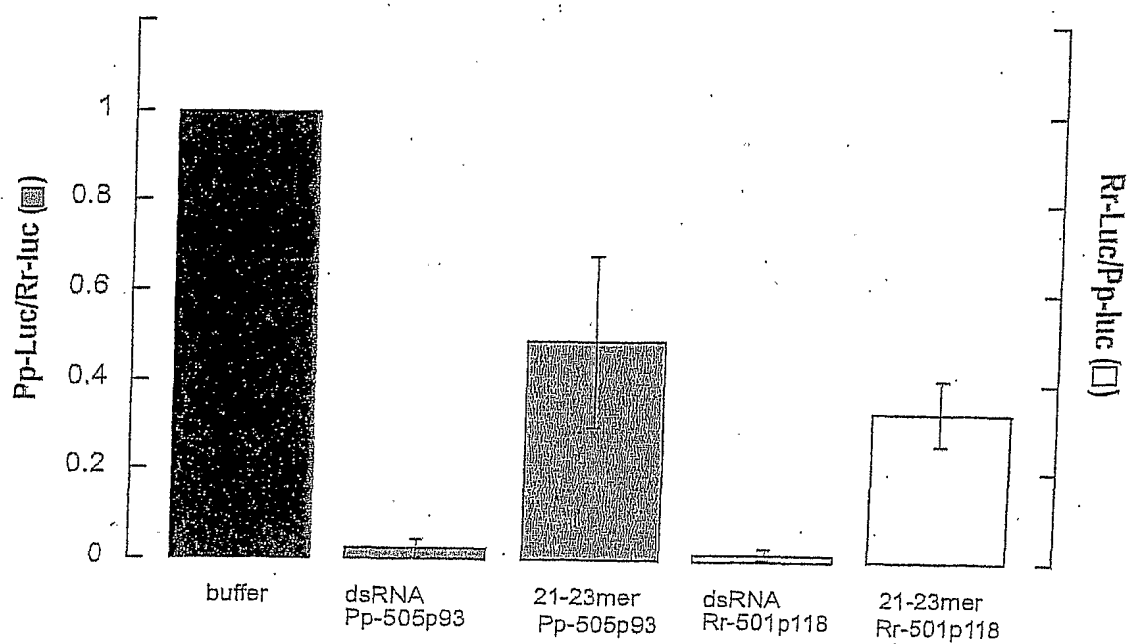
Figure 12
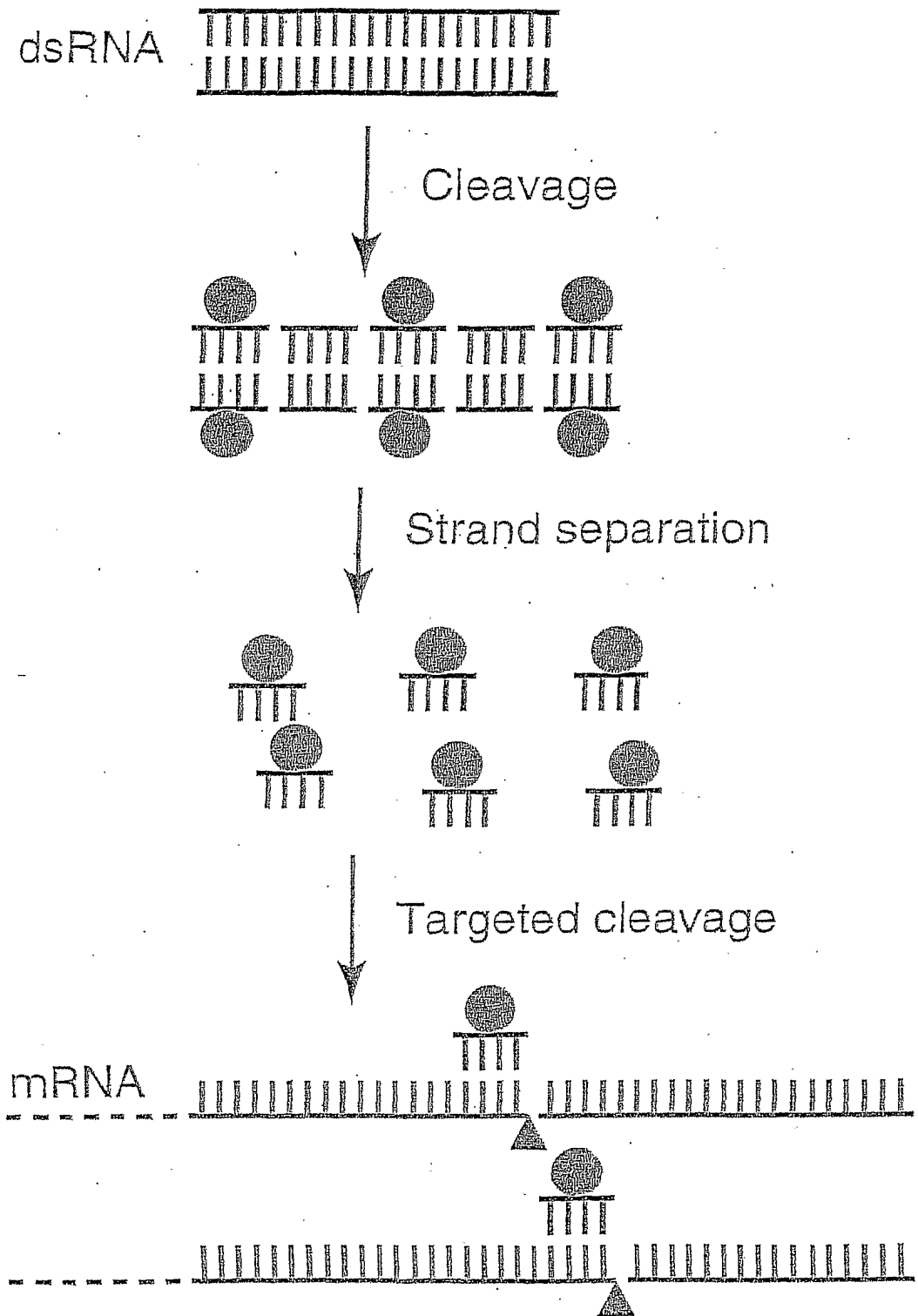
Figure 11
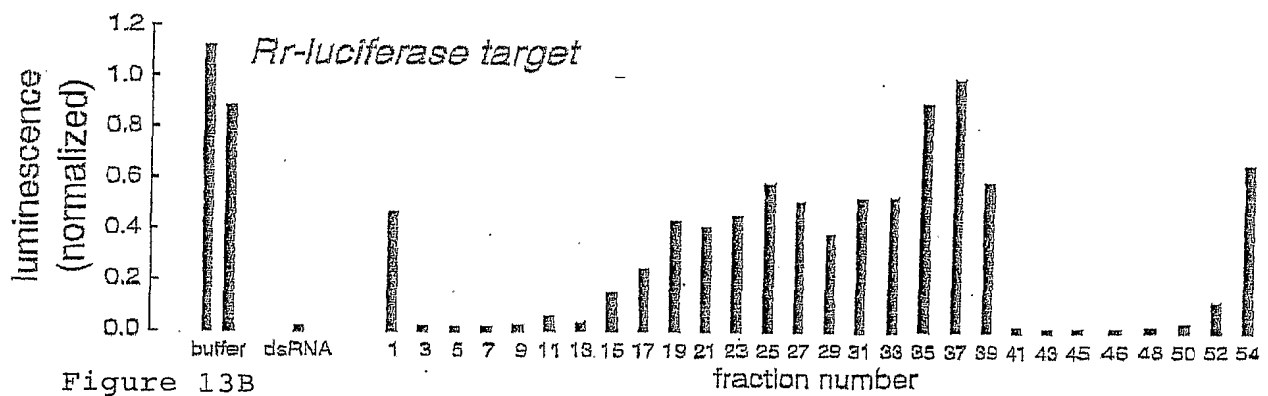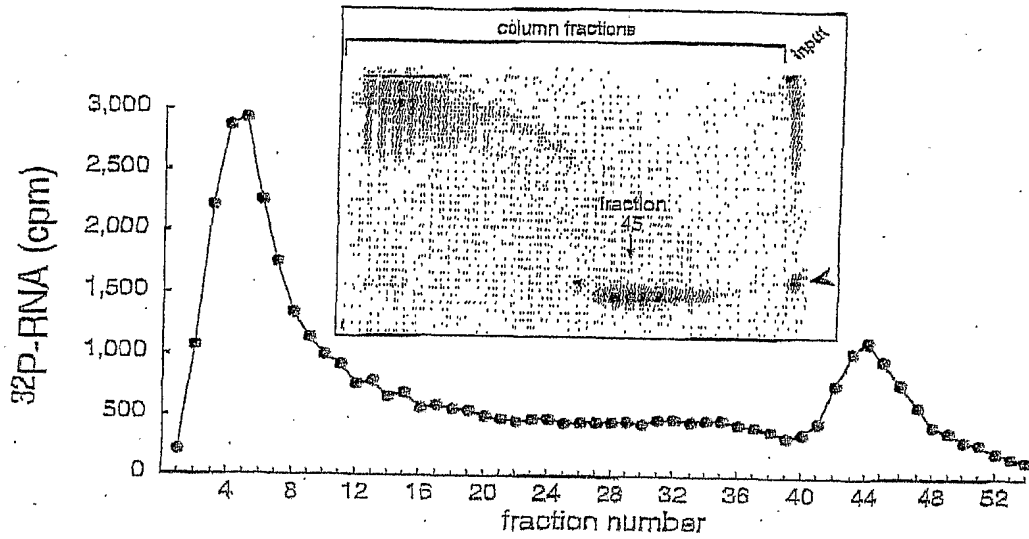
Figure 13A

Figure 13B
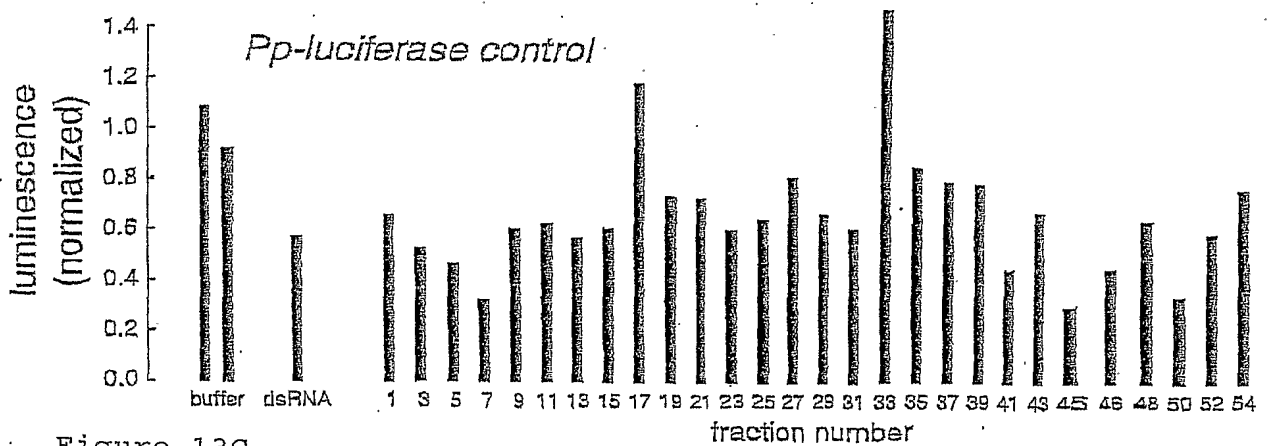
Figure 13C
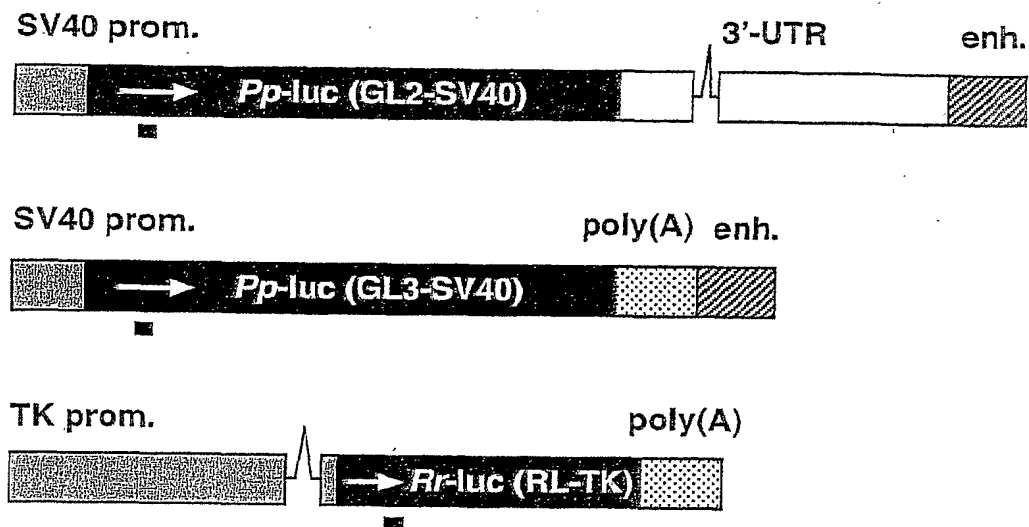
Figure 14A

siRNA
duplex

uGL2   5′ CGUACGCGGAAUACUUCGAUU
            UUGCAUGCGCCUUAUGAAGCU 5′

GL2   5′ CGUACGCGGAAUACUUCGATT
            TTGCAUGCGCCUUAUGAAGCU 5′

GL3   5′ CUUACGCUGAGUACUUCGATT
            TTGAAUGCGACUCAUGAAGCU 5′

invGL2   5′ AGCUUCAUAAGGCGCAUGCTT
            TTUCGAAGUAUUCCGCGUACG 5′

RL   5′ AAACAUGCAGAAAAUGCUGTT
            TTUUUGUACGUCUUUUACGAC 5′

Figure 14B

Figure 15A

Figure 15B
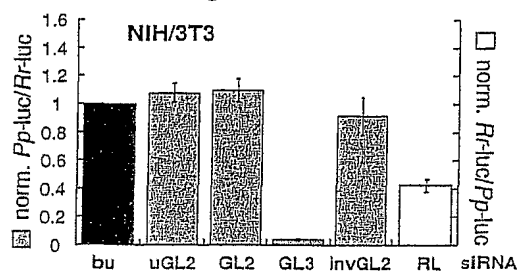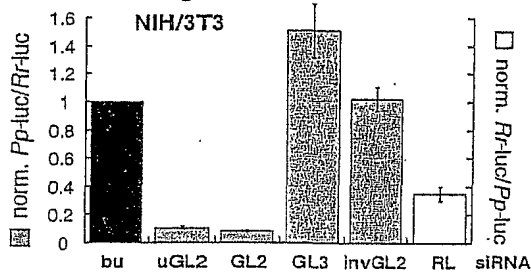
Figure 15C

Figure 15D
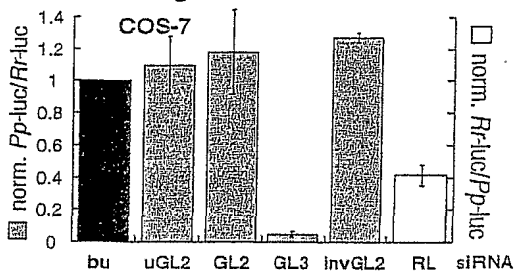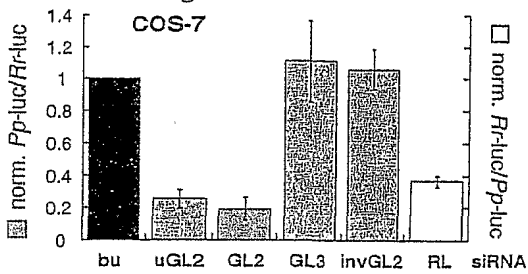
Figure 15E

Figure 15F
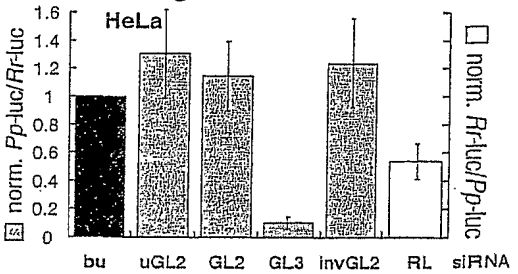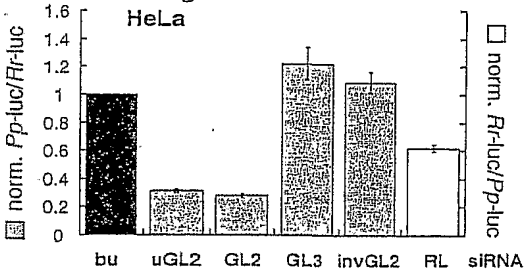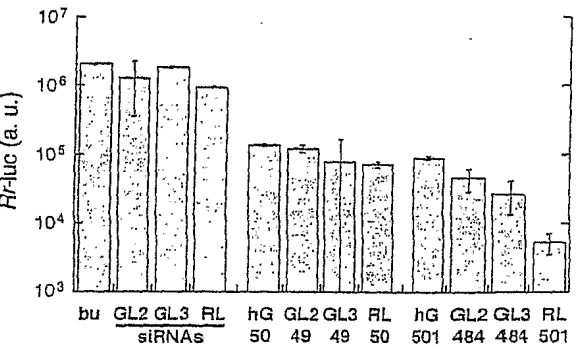
Figure 15G

Figure 15H
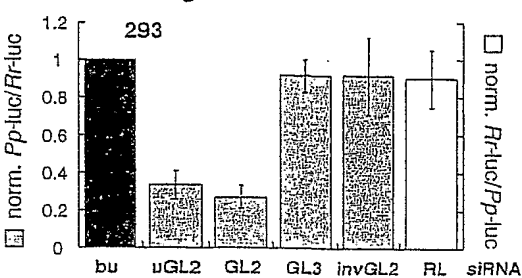
Figure 15I

Figure 15J
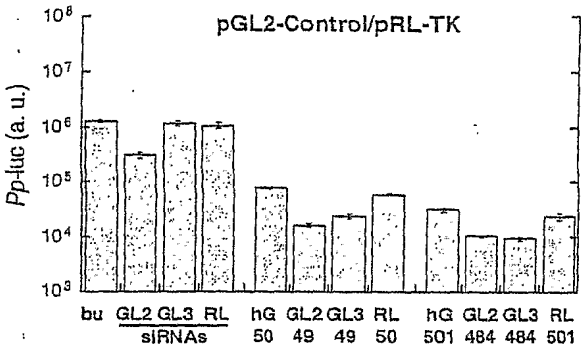
Figure 16A
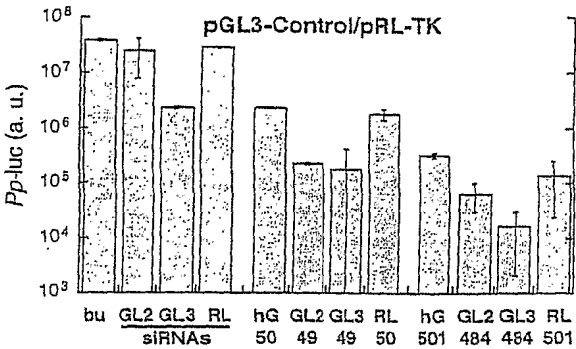
Figure 16B
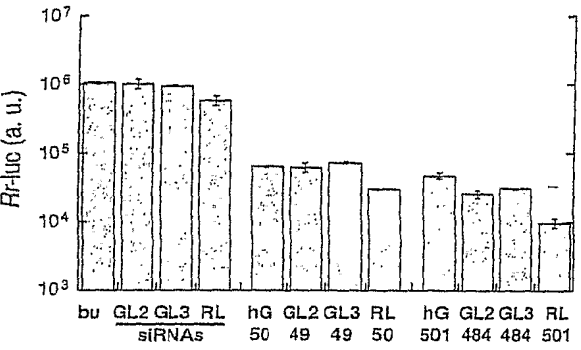
Figure 16C

Figure 16D

Figure 16E

Figure 16F